StoryBuddy: A Human-AI Collaborative Chatbot for Parent-Child Interactive Storytelling with Flexible Parental Involvement

Zheng Zhang*
zzhang37@nd.edu
University of Notre Dame
Notre Dame, IN, USA

Ying Xu*
ying.xu@uci.edu
University of California, Irvine
Irvine, CA, USA

Yanhao Wang
yanhaow@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

Bingsheng Yao
yaob@rpi.edu
Rensselaer Polytechnic Institute
Troy, NY, USA

Daniel Ritchie
drritchie@uci.edu
University of California, Irvine
Irvine, CA, USA

Tongshuang Wu
wtshuang@cs.washington.edu
University of Washington
Seattle, WA, USA

Mo Yu
yum@us.ibm.com
IBM Research
Yorktown Heights, NY, USA

Dakuo Wang[†]
dakuo.wang@ibm.com
IBM Research
Cambridge, MA, USA

Toby Jia-Jun Li[†]
toby.j.li@nd.edu
University of Notre Dame
Notre Dame, IN, USA

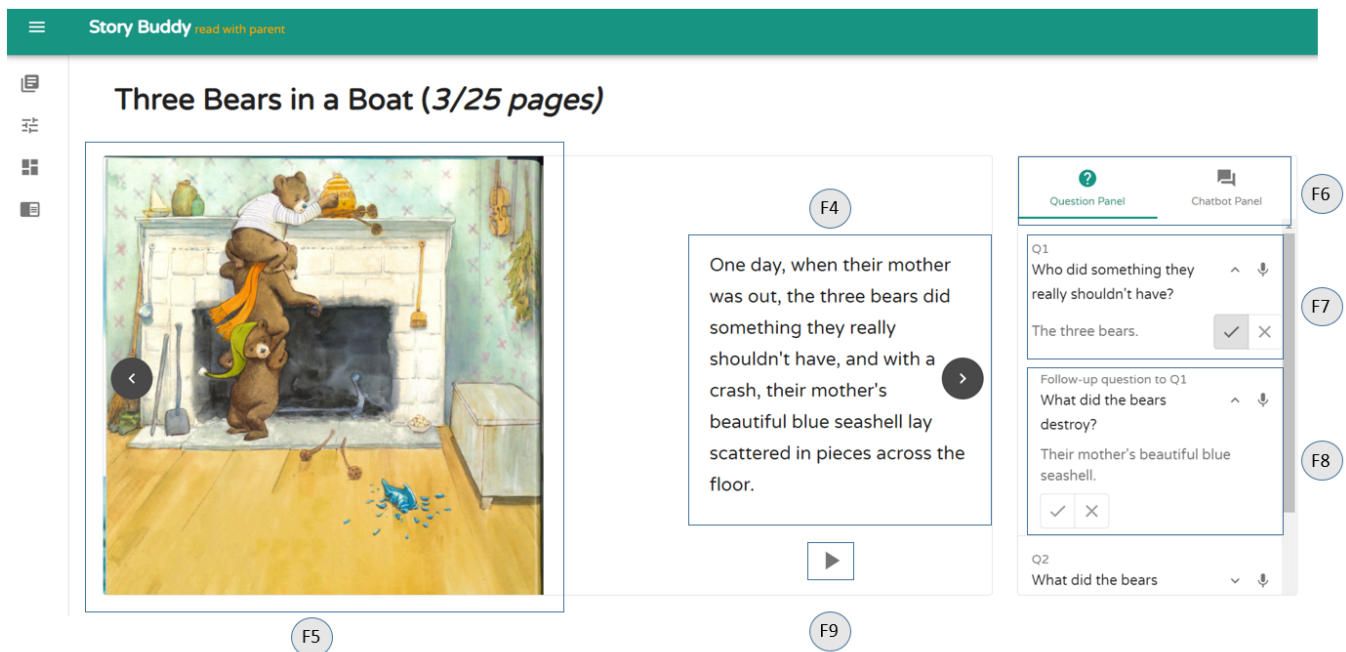**Figure 1: StoryBuddy's story reading interface in the parent-AI co-reading mode**

*equal contribution
[†]corresponding authors

ABSTRACT

Despite its benefits for children's skill development and parent-child bonding, many parents do not often engage in interactive storytelling by having story-related dialogues with their child due to limited availability or challenges in coming up with appropriate questions. While recent advances made AI generation of questions

from stories possible, the fully-automated approach excludes parent involvement, disregards educational goals, and underoptimizes for child engagement. Informed by need-finding interviews and participatory design (PD) results, we developed StoryBuddy, an AI-enabled system for parents to create interactive storytelling experiences. StoryBuddy's design highlighted the need for accommodating dynamic user needs between the desire for parent involvement and parent-child bonding and the goal of minimizing parent intervention when busy. The PD revealed varied assessment and educational goals of parents, which StoryBuddy addressed by supporting configuring question types and tracking child progress. A user study validated StoryBuddy's usability and suggested design insights for future parent-AI collaboration systems.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Natural language interfaces*; • **Social and professional topics → Children**.

KEYWORDS

interactive storytelling, co-reading, dialogic reading, voice user interfaces, human-AI collaboration, child-agent interactions

1 INTRODUCTION

Storytelling is a common parent-child activity that provides many educational benefits such as improving children's language fluency, communication skills, cultural and emotional awareness, and other aspects of cognitive development [60, 79]. Interactive storytelling in particular, where a storyteller asks questions relevant to story content and prompts a child to express their thoughts about the story, has been shown to maximize the educational benefits of storytelling [31]. Many parents experience barriers like difficulty in coming up with appropriate questions, high cognitive load from multi-tasking, and challenges with keeping track of the child's progress. To address these barriers, many digital systems, both from the industry and research community, have been proposed to facilitate parent-child interactive storytelling with children.

Many prior digital interactive storytelling systems have been shown to be effective in supporting various learning goals. For example, StoryCoder [12] leverages storytelling as a creative activity by allowing children to first listen to stories and then modify these stories in computational thinking games. This approach was shown to be effective in the development of computational thinking. Conversational agents were also developed to support children's literacy development [81], bilingual language acquisition [6], and foster science learning [80]. In the HCI community, several empirical studies have been done to investigate how children and parents interact with existing voice agents such as Amazon Alexa and

Google Home [4, 44, 78, 84, 86]. These studies identified opportunities in the use of voice agents to facilitate learning, development, and social goals of children, but also pointed out challenges in facilitating child-agent interaction. Commercial products such as Luka [43] and Codi [35] are AI-enabled robot toys that can facilitate interactive storytelling experiences. Codi is a storytelling robot that can tell over 100 pre-recorded stories. Luka is a "AI reading companion" that the child can place in front of a book while the child reads the book. Luka can recognize the book (from 20,000 books in the developer's library) and ask the child preset questions relevant to the story. An important limitation of all these existing systems for interactive storytelling is that their questions are manually crafted—therefore they only support a limited set of books or stories that the system developers prepared.

Recent advances in natural language comprehension and question generation (e.g., [15, 32, 66]) made it feasible to automatically generate question-answer pairs (QA pairs) about story plots from *any* storybooks, enabling fully automated interactive question-answering between children and a chatbot. But there are several issues in the adoption of this approach in real-life storytelling sessions with children:

(1) While the vast majority of generated QA pairs are syntactically and factually correct, many do not serve educational purposes (e.g., too trivial, not relevant to the main story plot) and are not necessarily appropriate (e.g., containing difficult words) for the children [24, 50].

(2) A fully automated approach excludes parent involvement—prior research shows that parent storytelling not only develops language and comprehension skills of the children but also strengthens the bond between parents and children [19, 74].

(3) Simply asking questions in sequence from a list of generated questions does not optimize for child engagement due to the lack of logical connection between questions and the lack of guidance for the children in case of confusion or incorrect answers.

To address the above limitations of an *AI-only* approach, we explore a *human-AI collaboration* approach that incorporates the expertise and preferences of parents into the development of interactive storytelling experiences for their children. However, there is no one-size-fits-all solution. Our formative study and participatory design process (Sections 3 and 4.1) found that parents have different motivations, objectives, and preferences for using digital storytelling systems, and they want a system that can adapt to various usage scenarios. This large variety of user needs results in diverse, sometimes even conflicting, design goals and constraints for the system. For example, parents reported that they regard storytelling as an important way to strengthen relationships between themselves and their children, therefore it is important for any AI assistance in storytelling to preserve direct parent-child interactions. At the same time, they expressed the desire for an automated storytelling system that can keep their children engaged without any parent intervention in situations where they need to focus on something else (e.g., when they are in a meeting while working from home). We also heard different opinions from parents on whether they prefer to put a stronger emphasis on the skill development and

assessment objectives in storytelling or if they wish to just "keep it fun" for their children as a form of entertainment.

Building on prior literature and our own formative investigation with 10 families, we designed and developed StoryBuddy, a new system that allows parents to collaborate with AI in creating storytelling experiences with interactive questioning-answering. Through co-design sessions with four parents using storyboards, we proposed an interaction strategy that supports two distinct modes: (1) an assisted *parent-AI co-reading mode* where the AI assists the parent in storytelling by identifying potential opportunities for asking questions and recommending follow-up questions. StoryBuddy in this mode can reduce the cognitive load and lower the literacy barrier for the parent, facilitate skill development for the child while encouraging direct parent-child interaction that both parties value in their relationship. (2) an asynchronous *automated bot-reading mode* where the parent can create an interactive storytelling bot for *any* story by configuring the question generation model, selecting from the generated questions, and customizing follow-up questions. The bot can then tell stories, ask children questions and provide feedback, and converse with the children to keep them engaged without intervention from the parents. StoryBuddy in both modes also tracks the child's progress and visualizes children's performance data in a dashboard, enabling the parent to assess the development of the child's comprehension skills.

This paper presents the following three main contributions:

- a formative interview study and a participatory design process with parents that uncover the large variations in parents' objectives for interactive storytelling, their need for the support of flexible parent involvement, their challenges with the high cognitive load from multitasking, and their desired strategies for enhancing child engagement in interactive storytelling.
- the design and implementation of StoryBuddy, a system where parents collaborate with AI in creating interactive storytelling experiences with question-answering for their children
- a user study with 12 pairs of parents and children that evaluates the usability of StoryBuddy and sheds light on how parents and children interact with StoryBuddy

From the findings of the design process, the implementation of the system artifact, and a user study with the system, this paper presents several implications for designing human-AI collaborative systems in facilitating parent-children interaction. Specifically, we (1) identified challenges in designing a workflow that accommodates effective *partial* automation in real-time that copes with interruptions and resumptions, (2) presented a design strategy of transforming synchronous involvement into asynchronous involvement to support flexibility in parent involvement, and (3) discussed opportunities in designing flexible multi-faceted roles of an AI companion that fits into the existing parent-child interaction dynamic in a familiar activity (storytelling) while balancing between multiple educational, developmental, assessment, engagement, and relationship-building goals.

2 RELATED WORK

2.1 Digital Systems for Facilitating Interactive Storytelling

Storytelling with children is an activity that provides significant benefits in skill development, relationship building, and entertainment [19, 60, 79]. Specifically, storytelling between parents and children is a routine activity in families across different cultures [67]. Interactive storytelling is a form of storytelling where, in addition to merely narrating the story verbatim, parents actively interact with their children about the story content. An effective and popular strategy in interactive storytelling is guided conversation [89] (also known as dialogic reading), where a storyteller asks a child questions about story content and provides responsive feedback. This strategy allows children to actively participate in the storytelling process, reflect on their comprehension of the story, and express their understanding through multi-turn dialogues. Prior studies in this area found that guided conversation with question-answering has positive impacts on the development of language and literacy skills for children [17, 49].

While there is evidence on the benefits of dialogic reading for children across a broad age range, much of the research attention has been focused on younger children aged three to eight who are at the stage of "learning to read" and do not have fluent decoding skills [36, 49]. Dialogic reading is particularly suitable for this age group as this reading activity is typically carried out orally, thus allowing young children to fully allocate their cognitive resources on making sense of the text they hear [29]. Thereby, dialogic reading promotes the kind of oral language skills, including vocabulary and narrative comprehension [23], that are strongly linked to children's later reading skills and academic success as they move to the "reading to learn" stage after the age of eight [69]. Nevertheless, children aged three to eight span two different stages of the reading development [25, 55]: a *pre-reading stage* when children gain mastery over the sound structure of spoken language, make inferences of stories from pictures, and develop listening comprehension skills; and an *early-reading stage* when children learn to decode print text and begin to read fluently and strategically. Due to pre-reading children's limited decoding skills, they primarily rely on stories being read to them, while children in the early-reading stage start to read stories more independently. Our system provides audio narration to support pre-reading children, but also allows early-reading children to read without audio narration. We will discuss the different usage patterns between these two stages in the findings of our user study (Section 6.3.1) and how future versions of StoryBuddy can better accommodate the specific needs of children in these age groups (Section 8).

Prior research has proposed several effective questioning strategies for dialogic reading. In general, there is a consensus that open-ended "Wh-" questions are more effective in eliciting children's verbal responses than yes-or-no or multiple-choice questions [59]. Furthermore, "Wh-"questions can be categorized based on the information required for formulating answers. Rubegni and colleagues suggested incorporating two types of "wh-"questions: basic prompts that focus on children's *recall* of story events and contexts, and "Theory of Mind" prompts that encourage children to *make inferences* of story characters' thoughts, feelings, and intentions [62].

Blewitt's study also observed the benefits of including both types of prompts [8]. Thus, in our research, we follow these evidence-based suggestions to incorporate both recall and inferential questions in our question-generation model.

Various digital tools have been introduced to help facilitate different aspects of interactive storytelling [82]. For example, Kory and Breazeal created an embodied learning companion robot that can introduce new vocabulary words during storytelling [30]. Michaelis and Mutlu designed an in-home learning companion robot that can make preprogrammed comments in stories [48]. StoryCoder presents two storytelling games and four computational thinking games to support the use of interactive storytelling as a way to teach computational thinking concepts [12]. There were also tools for supporting the creation of multimedia stories such as Fiabot [63] and StoryBank [18]. Besides academic research work, commercial products such as Luka [43] and Codi [35] also support automated telling of stories and child-bot conversation about the story content.

Compared with prior work, the two key novel contributions of StoryBuddy are (1) its support for interactive question-answering on *any* stories; and (2) its new design features for supporting flexible levels of parent involvement. The systems we discussed above rely on manually prepared story-specific questions. In comparison, enabled by a state-of-art question-answer generation model, Story-Buddy can automatically generate appropriate questions, identify follow-up questions, and engage in multi-turn question-answering with children for any stories. Several parent-AI collaborative mechanisms in StoryBuddy help parents ensure that the generated questions (1) are appropriate for their child; and (2) can serve the intended goals parents have. Previous systems also lack support for flexible parent involvement, which is a key user need according to both prior literature [41] and our formative study findings. Story-Buddy's two distinct modes support situations for both when the parent is present and when the parent is absent. In StoryBuddy, parents who wish to have more control have the option to customize the question-answering content, select generated question types, and track child progress. For others, these steps can be automated with little parental intervention.

2.2 Studies of Child-Agent Interaction

As we discussed before, dialogic reading resolves around back-and-forth conversation between adults and children. This makes conversational agents favorably positioned to act as children's reading partners. This technology has the affordances to understand unconstrained natural language input, thus allowing for complex dialogue and potentially mimicking human-to-human spoken conversation. Researchers recognized that conversational technologies can potentially offer a potent new mechanism for teaching, engaging, and supporting children in daily life [20]. The resulting developments may be especially valuable for young children, as their lack of proficiency in reading and writing cause difficulty for them to navigate many digital contents.

The design of StoryBuddy was informed by not only the results from our formative study and participatory design process but also insights from prior studies on how children interact with conversational agents.

Prior work identified opportunities in the use of conversational bots for facilitating learning, development, and social goals of children [4, 16, 21, 47, 81, 83, 85]. For example, a study by Beneteau et al. investigated how parents and children interact with Amazon Alexa in family homes through a 4-week deployment study. The study results suggested that the use of voice interfaces naturally promotes verbal communication and expands the communication skills of children. Parents found opportunities to use conversational bots to augment parenting practices. Specifically, they can complement parenting tasks (including storytelling) and increase the autonomy of their children [4]. A co-design study by Garg and Sengupta suggested that a conversational agent can be an ideal learning companion for children. Especially, parents want these agents to include them in the learning activities and to allow them to monitor their children's use [21]. Voice interfaces were also found to be effective in keeping children engaged [57]. Findings from these studies motivated StoryBuddy's design strategies in child skill development and assessment through conversational question-answering, enhancing parental involvement and customizability, and supporting a dynamic interaction paradigm that combines parent-child interaction with child-agent interaction to improve child engagement.

Prior studies also identified challenges specifically in facilitating child-agent interaction. While communication breakdown is common in general human-agent conversation [5, 11, 38, 46], and mechanisms such as [2, 39, 52] have been proposed to handle them, children's limited communication skills make it more difficult to avoid or repair communication breakdowns in child-agent interaction—Children are likely to encounter difficulties in understanding instructions, fail to follow the conversation flow, and struggle with appropriate turn-taking when interacting with agents [16, 45, 57]. In StoryBuddy, parents are involved in the child-agent interaction in the parent-AI co-reading mode, which alleviates these challenges by allowing the parent to help with breakdown repair. In both modes of StoryBuddy, controls in the graphical user interface (GUI) are available alongside the voice interface, so that the parent or child can still proceed through multi-modal interaction [56] in conversation breakdown situations [40].

2.3 Systems for Question-Answer Generation

StoryBuddy belongs to a category of systems that automatically generate questions and answers for a given piece of text (known as QAG systems in the natural language processing (NLP) community), but StoryBuddy's design goals and intended context of use are quite different from the vast majority of existing QAG systems. Most works on QAG systems approach the problem from a pure machine learning perspective, trying to invent new rule-based (e.g., [32, 88]) or neural-network-based (e.g., [14, 15, 65, 73, 77]) models that can generate "more accurate" questions and answers. This "accuracy" is commonly measured using objective similarity-based metrics (e.g., BLEU [58] which measures the precision of n-grams, ROGUE [42] which measures the recall of n-grams) that compare the generated questions against the gold standard of human-generated questions. While these systems perform well in generating *correct* and *relevant* questions. Their generated questions usually lack educational values and are ineffective in maintaining child engagement, because these QAG systems are not optimized for these objectives.

Unlike prior systems, the question-answer generation model used in StoryBuddy was specifically designed to be "as if a teacher or parent is to think of a question to improve children's language comprehension ability while reading a story to them [87]." and was trained on a dataset of children's storybooks annotated by educational experts for supporting interactive storytelling. Another category of relevant work is on interactive question-answering systems that seek to retrieve answers to questions that users ask (e.g., [61, 70]). These systems focus on answer retrieval (instead of question generation) and therefore have quite different goals from our work.

Compared with prior work, StoryBuddy emphasizes interaction design and human-AI collaboration aspects. In StoryBuddy, parents are heavily involved in the pre-configuration, question selection, and follow-up question generation process to better adapt AI-generated questions to interactive storytelling. In comparison, most existing work in QAG only focuses on the model without considering the intended context of use and the goals of the users.

3 FORMATIVE STUDY

As discussed above, there is a range of systems, tools, and applications aiming to support storytelling in early childhood. However, these solutions have not been designed to promote parent involvement or provide personalized reading experiences for individual families. This is less ideal for supporting children and families' diverse needs. To further understand this issue from the users' perspectives, we conducted a formative study to gather information on (1) families' daily practices of storytelling and digital device usage and (2) families' general needs and expectations of digital storytelling systems.

3.1 Method

We recruited and interviewed ten families with at least one child aged three to eight years from two different communities in the Western U.S., including one predominately White and Asian University community and one nearby working-class, Spanish-speaking community. Detailed participant information is displayed in Table 1. As shown in the table, all participants on average spent sometime on storybook reading daily. The "Parent-Child Language Use" column indicates the language used at home between parents and children. Each interview session followed a semi-structured format, in which we asked parents questions about how they used digital media and devices for storybook reading with their children. We also asked parents about their general attitudes toward existing storytelling applications and their suggestions for improving these applications. We purposely started the interview with broad questions that placed fewer restrictions on the participants' responses and then asked more focused follow-up questions to probe parents' elaborations on certain topics. The interviews lasted 60 minutes for each participant and were carried out via video conferencing.

We used an inductive process to analyze the interviews. We began with qualitative memoing [7], in which members of the research team viewed the same portion of the data together, with each researcher individually memoing their own notes. After specific intervals (usually 5 minutes) researchers would pause data playback and discuss with one another the meaning that emerged from the data. During this process, we noticed emerging themes related to

parents' perception and expectation of digital reading technologies. We then systematically coded all the interview transcriptions based on the emerging themes, developing and revising codes as we found additional themes of parent perception and expectation. Coding was periodically cross-checked by two coders to ensure accuracy.

3.2 Key Insights

KI1: Parents value the educational affordance of new technologies. Nine out of the ten parents in our sample recognized that digital technologies and AI can support their children's language and literacy learning by improving letter recognition, phoneme awareness, vocabulary, spelling, and story comprehension. Though popular press often implies that busy parents use digital devices simply as a "babysitting" tool, our interview indicated that parents intentionally use technologies as enriching educational opportunities for their children, especially when it concerns domains that they do not think they are capable of teaching. For example, one parent in our study said " *(translated from Spanish) I don't speak English at all. So letting my little one watch television or use apps are important for her to learn English before school.*" Another non-English speaking parent showed us a Luka device in their home - an AI-powered robot that can read print books aloud –commenting that they used this device as their young child's "English learning time." A parent who does not speak Spanish herself mentioned that her child sometimes picked up Spanish words from talking with Alexa. The parent said "*It's just like she can have this fun Spanish lesson with Alexa. I think, wow, that is cool.*" Moreover, affordable digital content is a valuable learning resource for families with less access to expensive educational opportunities, such as private tutoring and enrichment camps. One parent mentioned that their child frequently played free spelling games on PBS KIDS, because *"it is just available for us"*.

KI2: Parents prefer interactive storytelling systems. Eight parents mentioned that they prefer technologies that provide interactive opportunities for children, which the parents believe lead to more engaging and active learning. For example, one parent viewed positively the choose-your-own-adventure story apps on Alexa that allowed her child to control how the story proceeds by providing speech command, pointing out that *"Compared to just listening to a story…I think the interactive one [can] get her brain thinking, get her brain building, and working, and exercising"*.

KI3: Parents view technologies as a way to promote parent-child interaction. While there is a fear that children's use of technologies supplants their interaction with other family members, Six parents in our study suggested that they think technologies can have a positive impact on family interactions, even though many of the technologies are not intentionally designed to encourage parent-child interaction. Four parents mentioned the enjoyable moments they have had watching television or playing video games with their child, which made them "*feel closer [with their child]*" or *"like doing a family thing"*. Another parent, in particular, mentioned that some interactive features of digital books (e.g., hotspots) often triggered their child to ask her questions or make comments to her, which often

ID	Parent Gender	Child Age	Parent-Child Language Use	Minutes Reading (Weekday)	Minutes Reading (Weekend)
BF1	Female	4	English and Spanish	30	0
BF2	Female	4	Mostly Spanish	30	10
BF3	Female	4	English and Spanish	5–10	0
BF4	Female	6	Mostly English	10–15	20
BF5	Female	5	Only Spanish	10	10–15
BF6	Female	4	Mostly Spanish	90	30
BF7	Female	7	Mostly Spanish	30	20
BF8	Female	5	Mostly English	30	0
BF9	Female	4	English and Spanish	20	0
BF10	Female	5	English and Spanish	20	0

Table 1: Demographics of formative study participants and the average daily time they spent on storybook reading.

turned into an interesting longer conversation. Nevertheless, parents do appreciate apps or systems that are intended for co-use, particularly those designed for educational purposes. One parent mentioned the challenges she had trying to interact with her child when using a story reading app. Although this parent was well aware of the benefits of asking children questions during reading, she told us that *"I'm trying to think of what questions I could ask my kid but sometimes I can't even think of anything off the top of my head"*.

KI4: *Parents are highly involved in selecting content for their children and have a desire for customized content.* We found that parents tend to carefully select digital content they think is beneficial for their children, as this theme was brought up by over half of the parents we interviewed. They either rely on their own subjective judgment (e.g., *"I just think it's good for my kid."*) or seek out guidelines issued by researchers or institutions. For example, a parent mentioned that *"I just saw this on Common Sense Media, which is where I usually go to have an initial check on age level appropriateness"*. Nevertheless, all parents expressed confidence that they know what is appropriate for their children because they know *"what [their child] likes, what [their child] knows, what [their child] doesn't like."* Therefore, it is not surprising that some parents indicated that they sometimes wish they could modify the content to better fit their child's interests or needs. For example, one parent mentioned *"yeah those are good apps, but I might want to change the language a bit. I don't think my son understands this word"*.

4 PARTICIPATORY DESIGN

4.1 Process

With the design opportunities identified in the formative study, we launched participatory design (PD) sessions [51] with four parents (PA1–4) to further uncover concrete design goals and design strategies for StoryBuddy.

4.1.1 Participants. The participants were parents recruited through the mailing lists of our maintained participant pool. Three of them spoke English as a second language but were fluent in English. Two of them were mothers of 4-year-old children and another two were mothers of children older than 5. All participants told stories frequently to their kids: Two parents told stories to their

kids once a day, while another two did storytelling 4–6 times per week. All of them were primary caregivers to their kids.

4.1.2 Procedure. The sessions were conducted remotely via Zoom and each lasted around an hour. In each session, we presented the parent with four sets of low-fidelity storyboards, each illustrating a different scenario of using StoryBuddy to tell stories to their kids. The parents were asked to discuss their feelings about the scenarios in the storyboards, identify possible design opportunities, challenges, and user concerns, and ideate design strategies and new features to improve the system. Each parent was compensated with a $25 gift card for their time.

The main goal of the PD approach is to include parents' voices and ideas in the design process, utilizing their unique experiences in helping us explore the problem space. Prior to each PD session, we conducted a semi-structured interview with the parent to learn more about their current practice and strategies in storytelling, whether they engage in any questioning-and-answering with their children, and any challenges they encountered in storytelling.

In each PD session, we presented the participant with four variations of storyboards in random order, an example storyboard is shown in Figure 2. Those storyboards serve as starting points in the PD process so that participants can brainstorm new interfaces, interaction strategies, use scenarios, and system capabilities based on the variations of user needs, interaction modalities, and contexts depicted in four storyboards [22, 51]. The four variations of storyboards were designed based on the key insights (Section 3.2) from the formative study as we will articulate below.

4.1.3 Storyboards. The four variations of storyboards differed from each other in two aspects: (1) whether the parent is present at the storytelling; and (2) whether the system runs on a tablet or a smart speaker. These two aspects reflected the options in key design decisions we identified from the formative study. The storyboards also showcased the envisioned user needs and corresponding system features on interactive conversational agents, the configuration of question types, the dashboard that supports child performance tracking, and the generation of follow-up questions.

The design of storyboards was based on Personas [22] constructed using the key insights from the formative study. In particular, the parent profile in the storyboards was characterized as
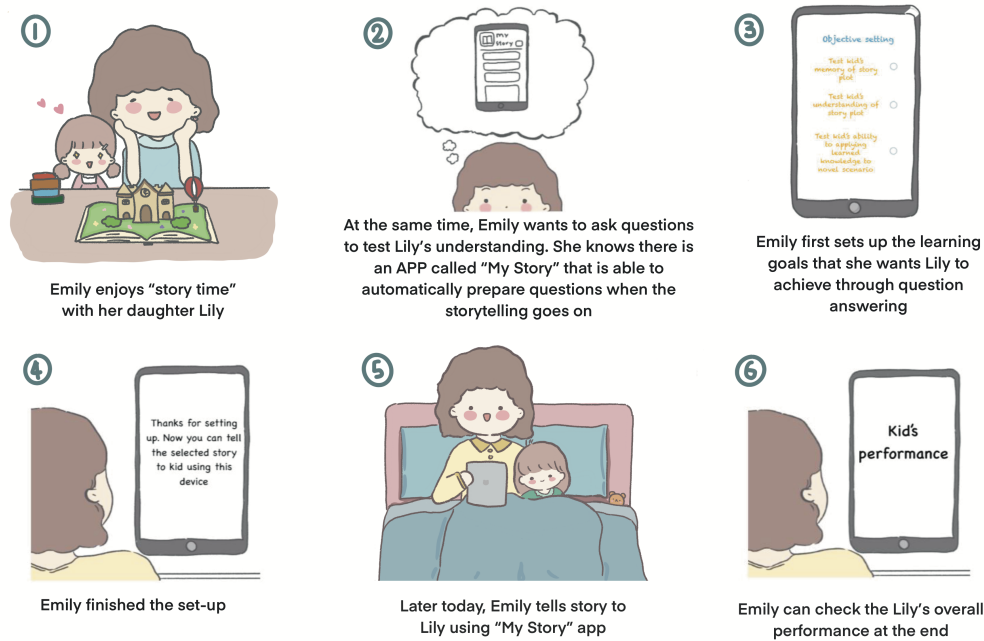
Figure 2: An example part of a storyboard used in our participatory design sessions

someone who appreciates the educational benefits in technologies for children (KI1), is willing to adopt new AI-enabled interactive systems in storytelling (KI2), values human-human interaction between parents and children (KI3), and wishes to stay involved in curating and filtering digital contents that their children consume (KI4). The parent was also sometimes "busy and unavailable" and therefore absent from the synchronous storytelling session in the "parent absent" variations of storyboards. The use of persona as a PD tool helps users understand the design challenges and concretize their ideas [22], which has also been previously used in similar design domains [1, 64, 71].

The storyboards set scenarios for users to reflect on their needs, constraints, and practices [9]. In the storyboards, we purposely deemphasized the details in the interface designs of the system by avoiding directly showing screen contents (if showing screens was necessary, our storyboards used low-fidelity sketches). Instead, the storyboards focused on illustrating the parent motivations and goals in the scenarios, the constraints in time, attention, and cognitive capabilities, and the interaction dynamic among the parent, child, and agent in different scenarios depending on whether the parent is present and which type of device is used. The main goal was to elicit the feelings and emotions of participants towards different design decisions and to build empathy with them. After validating user needs illustrated in the storyboards, we asked participants to think about how their personal experience with parent-child storytelling (or the lack of it) can connect to the storyboard scenarios. Through this process, each parent identified things they liked, things they disliked, and their concerns about the different paradigms of AI involvement in the parent-child interactive storytelling process. Lastly, we asked them to think of and propose new ideas on design

features, interfaces, or interaction techniques for (1) addressing the issues they identified in the storyboards; and (2) bridging the gaps between the scenarios presented in the storyboards and their own personal scenarios.

4.2 Findings

4.2.1 Optimizing for engagement as a key goal in storytelling. Achieving high child engagement in storytelling is a key goal for parents in both kinds of scenarios: when the parent is present and when the parent is absent. When a parent participates in the storytelling, maintaining child attention is challenging as children quickly get bored and subsequently distracted when there is a lack of change in the type of activities or interaction patterns. This issue becomes more problematic when the parent is not present. When encountering time and attention conflicts, parents often seek to use digital content such as videos of stories and songs on smartphones to keep children occupied when parents are in meetings (especially common during the COVID-19 pandemic when many parents work from home and have remote meetings) or doing housework. However, children quickly get bored and try to seek attention from their parents when they are unavailable, resulting in frustration for both parties (PA1, PA2).

When we asked about practical strategies that participants currently used to enhance children's engagement in storytelling (without the involvement of an intelligent agent), joint reading came up as a common strategy that parents usually adopted: "*Usually, we'll take turns since she doesn't like to read all the pages by herself. So, so I want to, like motivate her. And usually, I'll read one sentence and she will read the next sentence and we will take turns to read the whole book.*" (PA2), "*So we do what's called joint reading. So*

I'll read, and then I'll have him read. And then I'll read and we go back and forth." (PA3). In addition to enhancing engagement, some participants also valued the effect of joint reading on providing emotional support to their kids: "*in addition to engagement, I think, (kids will get) the emotional support, while you're reading with your child, and the interaction with the child, especially before (going to) bed, ... your child will feel very loved and warm*" (PA2).

The PD sessions identified providing multiple variations of interaction patterns as a key strategy for improving engagement. An ideal digital system for interactive storytelling should support (1) flexible switching between who is reading and who is facilitating the questions among the agent, the parent, and the child; (2) diverse question-answering patterns with various question types, follow-up questions on correct answers, and guidance for incorrect answers.

An ideal digital storytelling system should also balance the parent's desire for involvement (as reported in Section 3) and their practical need for minimizing parental intervention needed when busy by providing distinct modes. In a parent-agent co-reading mode, the agent should play a *supportive* role that helps the parent identify opportunities for questions, recommends appropriate questions, and offers options for occasional child-bot interaction to make the process more "fun" and engaging for the child. While in a parent-absence mode, the agent needs to play a *proactive* role in engaging the child with the goal of minimizing their need for parental attention. Nevertheless, despite the lack of *synchronous* parent involvement at the time of storytelling in this mode, many parents still desire *asynchronous* parent involvement through configuring the agent's interaction plan beforehand and tracking the child's progress afterward.

4.2.2 Challenges and opportunities of question-answering in digital storytelling. In current joint-reading practice, all four participants used questioning-and-answering to (1) improve engagement with children; and (2) assess and develop their comprehension in the storytelling. The participants often asked simple questions whose answers were apparent in the story, such as simple math questions (PA1), questions about color or shade (PA1, PA4), questions about pictures (PA2), or questions about major actions in the story (PA3). However, participants also recognized the importance of asking questions of different varieties and reported that children might have different demands for questions as they grow up: "*I think it (question type) depends on age, like, when kids are smaller, ... they want someone to ask very specific questions. And maybe for other kids (who are older), they want more challenging questions, or they don't want someone to interrupt them during the reading, instead, they prefer to answer the question afterward or before reading*" (PA2)

Coming up with appropriate questions of different types, especially in real-time while reading the story and interacting with the child can be challenging. As reported by PA4, sometimes she ends up finishing reading a story without asking questions due to the limitation in her cognitive load despite knowing about the benefits of question-answering.

4.2.3 Varied assessment goals of parents in question-answering. Among four participants, PA3 was the only one who explicitly reported using question-answering as an assessment tool ("*to make sure they are understanding the facts*"). She often liked to her child

follow-up questions on rationale (e.g., why is that?) and emotion (e.g., how do you think Susie will feel?) after questions about facts in the story plots (e.g., what happened...?) in order to assess the development of different skills of her child. Other parents also reported asking follow-up questions but as a means for maintaining child engagement instead of assessment.

Parents liked the idea of using AI assistance for identifying opportunities of asking questions and generating possible questions to use during parent-child joint reading. In the PD process, both PA2 and PA3 recommended design strategies for grouping similar questions together by themes, relevant entities, or question types to make it easier for parents to plan for them.

However, parents had diverging opinions on the assessment-focused features of the system. While PA3 was quite excited about the idea and suggested how assessment goals can be grouped into testing specific capabilities of the child, PA4 was concerned about whether the assessment goals embedded in the generated questions would affect the child's interest in the storytelling agent. She expected that the child would dislike the agent once they discovered that the agent was trying to assess them.

4.2.4 Balancing between desires and constraints in the granularity and form of parental involvement. Results from our PD sessions confirmed findings from the formative study that parents wish to have active involvement in the selection and configuration of contents in AI-facilitated storytelling, but the desired degree of involvement varies. PA1, PA2, and PA4 were OK with configuring question types, but were not too keen on the idea of editing and tweaking each individual question recommended by the system. However, PA3 reported her desire for controlling questions at a fine level of granularity despite that she did not expect most other parents to like it. PA3 said "*I think this [configuring individual questions] will overwhelm parents...[parents] want to rely on the app to figure out what those questions are. And you know, that you don't want to have to think about it...I would definitely love this, I think it could be an option.*" To address this divergence in user needs, our system should support flexible levels of parent control in story and question contents.

When it comes to the parent involvement in the delivery of contents, a conflict arises between (1) the parent's desire to be present, play an active role and foster the parent-child relationship; and (2) the constraint that sometimes they are not available is present for all four participants, which is consistent with the findings in the formative study (Section 3). Parents liked the fact that two variants of the system described in the storyboard can facilitate the storytelling process and the interaction with the child on its own without parent intervention, while still allowing asynchronous involvement from parents by controlling the story and question content beforehand and tracking the child's progress through a dashboard afterward.

4.3 Design Strategies

The PD process helped us identify the following six key design strategies, which will guide the design of the system described in Section 5:

- **DS1**: Maintain child attention and optimize for child engagement through the alternating use of many variations

of interaction means and approaches among the parent, the child, and the storytelling agent (Section 4.2.1).

- **DS2**: Assist parents with facilitating question-answering in joint-reading through recommendations of questions to ask and opportunities to ask questions (Section 4.2.2).
- **DS3**: Support different parental goals on whether to focus on the assessment objective in interactive storytelling (Section 4.2.3).
- **DS4**: Generate appropriate follow-up questions with relevant semantic entities or story themes to improve child engagement and the effectiveness of assessments (Section 4.2.3).
- **DS5**: Accommodate varied parent preferences on the granularity to configure interaction contents through providing multiple flexible options (Section 4.2.4).
- **DS6**: Address the conflict between (1) parents' desire to be present and involved in the live storytelling process in order to strengthen parent-child relationships and (2) the constraint that sometimes they are not available by supporting both (1) synchronous parent-child joint-reading with AI assistance and (2) AI-facilitated storytelling where question-answering contents were asynchronously configured by parents (Section 4.2.4).

5 STORYBUDDY

5.1 System Overview

Following these six design strategies, we designed and implemented StoryBuddy, an AI-enabled interactive tool for configuring, augmenting, and automating interactive storytelling with children. StoryBuddy presents several features that allow flexible parent involvement in both the configuration and delivery of interactive story contents, while supporting diverse parent needs in children's skill development, progress assessment, and engagement.

As shown in Figure 3, StoryBuddy consists of: (1) a storytelling configuration interface for parents to configure the question answering contents; (2) a parent-child co-reading interface for assisting the parent with the joint-reading process; (3) a conversational agent that can coordinate question-answering and automate storytelling when the parent is absent; (4) a dashboard that tracks and displays the child's progress; and (5) a back-end machine learning model that can generate possible questions and answers for *any* story.

5.2 Modes for Parent Presence and Absence

Informed by design strategies DS3, DS5, and DS6 from the formative study and PD sessions, we decided to create two distinct modes in StoryBuddy to reconcile the parent's desire to be present, play an active role, and strengthen the parent-child relationship with the constraint that they are sometimes not available for live storytelling.

5.2.1 Parent presence. As identified in the PD sessions, when the parent is present, the main goal of StoryBuddy is to assist the parent by helping them identify opportunities to ask questions and recommend questions to use in order to help reduce their cognitive load (DS2), while at the same time augmenting the parent by providing new variations of interactions for the child so they can interact with a conversational bot *in addition* to their parent to enhance

the child's engagement (DS1). It is important that StoryBuddy does not *displace* or *lessen* the parent's role, preserving the parent-child relationship-building aspect of storytelling that both the parent and child treasure (DS6).

To use StoryBuddy in the parent-child joint reading mode, the parent can simply choose a story from the story library panel (Figure 4). and enter the story reading interface (Figure 1). The story content panel on the left displays story text (F4) and the corresponding illustration (F5) of the current page. Parents can navigate through the pages by clicking left and right button.

On the right side, there is the question panel (F7 and F8 in Figure 1) that displays the AI-generated recommended questions, when the parent selects a question, the corresponding part of the story is highlighted, indicating the connection between questions and story contents. The parent can read the story to the child using the story content panel first, and decide if they want to ask any of the recommended questions from the question panel. They may click on the question so it expands to show the correct answer, and click on either the check or the cross button to record the correctness of the child's answer to be aggregated in the dashboard (see Section 5.4). Clicking on the check or the cross button triggers the generation of a follow-up question. The follow-up question will be about a relevant entity or a different aspect of the same entity in the original question (see Section 5.5.3).

An objective of our design of the parent-child joint reading mode is to give the parent a maximum level of control and agency (DS5). If they like, they could handle almost all aspects of the storytelling themselves without taking advantage of any "smart" features. However, they could also feel free to use the recommended questions and the generated follow-up questions as they see fit. They might also delegate question-asking, answer-checking, progress-tracking, and even the reading of the story itself to the agent as they wish (DS2). These features correspond to the PD finding that parents wish to have flexible degrees and granularity of control of their involvement in the digital storytelling experience.

Dynamic interaction paradigms for engagement. Maintaining child engagement is an important goal in parent-child joint reading mode. As identified in the PD, a potentially effective strategy is for StoryBuddy to support dynamic interaction paradigms so that the parent and the child can switch between different ways of interaction (DS1). This way, the child does not get easily bored. In parent-child joint reading mode, the parent can change the way of interaction on three aspects: (1) While the default setting is for the parent to read the story, the parent may easily have the agent read the story by clicking on the play icon, as shown in Figure 1. (2) After the parent asks a question, the child can either answer the question by speaking to the parent who will manually check the correctness of the answer using the question panel, or speaking to the agent by clicking on the microphone icon, as shown in Figure 1. When the child speaks to the agent, the agent can judge the correctness of the answer and further engage with the child (more details discussed in Section 5.2.2 about the automated bot-reading mode). (3) The parent may also quickly invoke the conversational agent to handle follow-up questions on their behalves in the parent-child joint reading mode by switching to the chatbot panel (F6 in Figure 1).
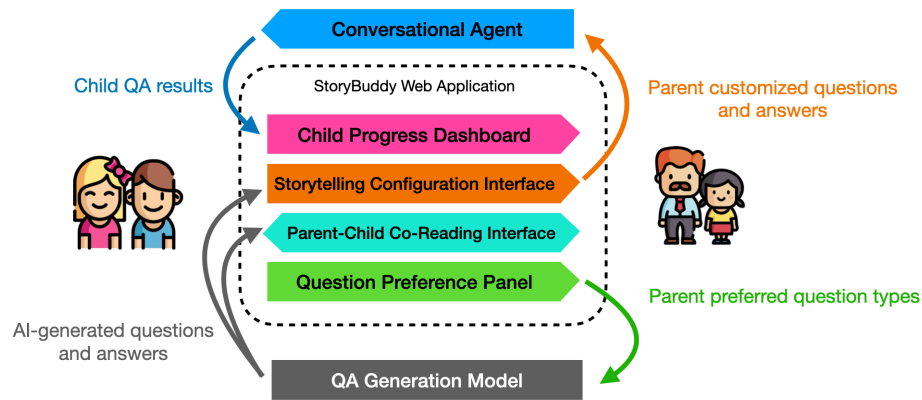
Figure 3: The system architecture of StoryBuddy. Lines with arrows represent data exchange between modules. The directions of the pointers on modules indicate whether the module interacts with the parents (right) or the children (left).
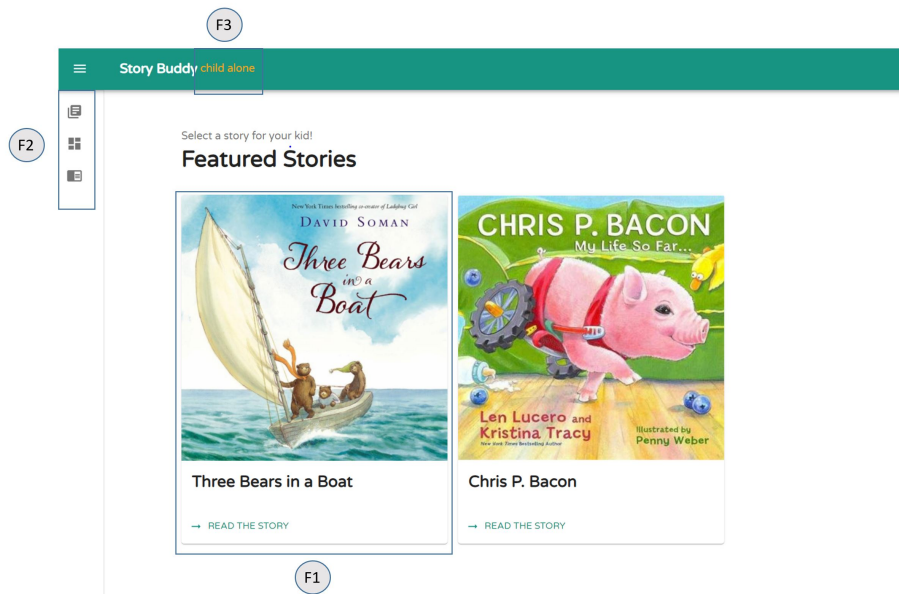

Figure 4: The story library panel in StoryBuddy

5.2.2 Parent absence. When the parent is absent, StoryBuddy operates in an automated bot-reading mode. The main goal of StoryBuddy in this mode is to engage children in interactive storytelling without requiring parents sitting next to the child during the storytelling (DS1, DS6). However, while parents are absent for the *synchronous* content delivery stage, often due to them being busy with other things, parents can still stay involved *asynchronously* through the configuration of the StoryBuddy agent and the tracking and assessment of child progress (DS4, DS5, dashboard details discussed in Section 5.4).

In this automated bot-reading mode, parents can configure how the StoryBuddy agent interacts with their child in advance by customizing the questions inserted in the stories. Similar to the parent-child joint-reading mode, the parent selects a story from the story library panel (Figure 4). They then enter the configuration page (Figure 5). The configuration page looks similar to the reading page in the previous mode, except for in the question panel, the parent can see all AI-generated questions and their corresponding follow-up questions in a list. The parent can choose which questions should be asked by the agent using the checkbox. In addition, the parent can edit the AI-generated questions and answers by clicking on the pen icon. This configuration step is optional–parents may directly click on the "*Proceed to read the story*" option to skip the rest of the configuration process. By default, StoryBuddy selects the top-ranked AI-generated question and its follow-up question for each page of the storybook.
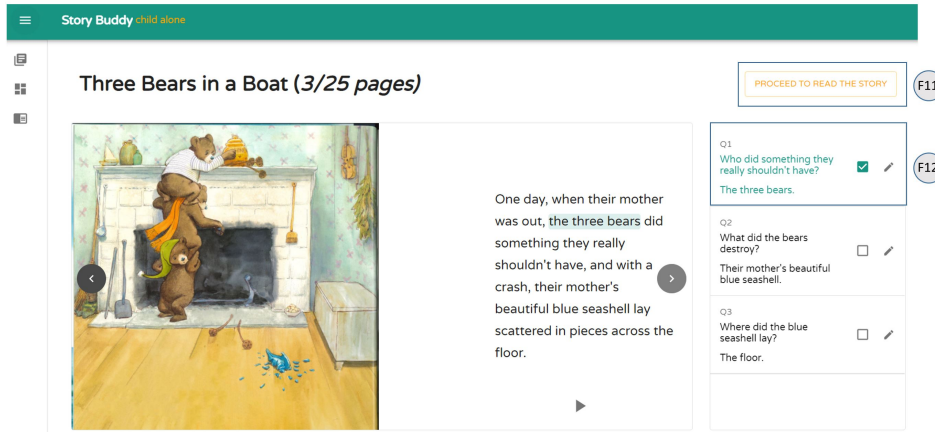
Figure 5: StoryBuddy's configuration page for questions to use in the automated bot-reading mode when the parent is absent
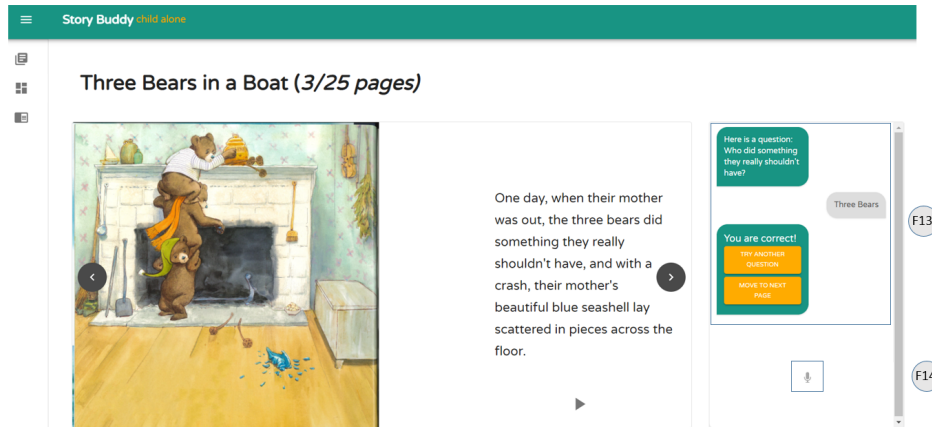

Figure 6: StoryBuddy's story reading interface in the automated bot-reading mode

Child-Agent Interaction in StoryBuddy. At storytelling time, children can see the book page including text and illustrations as shown in Figure 6. On the first page, the agent first greets the child. The agent will then read the story text on each page, say "*OK, here is a question*" and then ask the question as configured by the parent. The child can answer the question to the bot by clicking on the microphone icon (F14), the transcript of their speech will automatically appear in the dialog (F13). After receiving the child's answer, the agent will judge the correctness of the answer (technical detail in Section 5.5). If the answer is correct, the agent will say "*You are correct! Good job!*", the child can choose either "*move to next page*" or "*try another question*" (if another question is available); if the answer is wrong, the child will see an additional "*try again*" option for them to retry the same question.

5.3 Preference Configuration of Question Types

StoryBuddy provides a preference configuration panel (Figure 7), where the parent can choose the preferred types of generated questions for the back-end model. From the PD insights, we learned that some parents wish to have controls of the generated questions at a finer granularity (DS5). The preference configuration panel allows them to customize the generation of questions to better align with the learning and development goals they have for their children. The use of this panel is optional.

StoryBuddy allows the parent to indicate their preferences for questions focusing on seven different narrative elements, including questions about story *characters, setting, feeling, actions, causal relationships, outcomes, and predictions of future events* [59]. Character questions either start with "who" and ask the child to identify a character in the story, or ask the child to use information in the story to describe the character (e.g., "How did the man's daughter
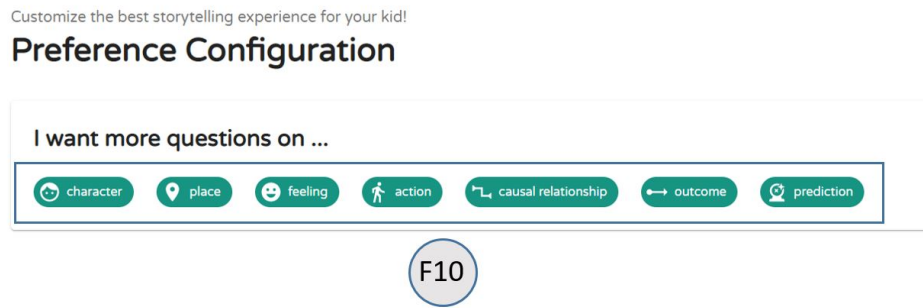
Figure 7: StoryBuddy's interface for question preferences configuration

look?"). Setting questions typically start with "where" or "when" and focus on a place or time that story events take place. Feeling questions ask the child to describe the emotion that characters experience (e.g., "How did the princess feel in her new home?") Action questions are typically phrased as "what does somebody do" or "how does somebody do something", for which the child needs to provide an answer that contains certain actions (e.g., "What did the cook do after she opened the hamper?" or "How did the prince break the curse on the princess?"). Causal relationship questions start with "Why" or "What makes..." that ask the child to identify the causes of a focal event in the story. Outcome questions ask the child to describe the outcomes or consequences of a focal event. Lastly, prediction questions ask the child to think about what might happen next (e.g., "How will the other animals treat the duckling?")

5.4 Dashboard for Child Progress Tracking and Performance Assessment

Another need of some (but not all) parents discovered from the PD insights is to track their child's progress and assess their child's performance (DS4). To address this, StoryBuddy provides an interactive dashboard (Figure 8). The dashboard can show either the child's performance in a particular individual storytelling session or the child's aggregated performance over a week. As shown in Figure 8, when the parent clicks on a previous storytelling session, the dashboard will show the parent information regarding each question that the child tried to answer in this session (F15), including the child's attempts and the right answer to each question. StoryBuddy also shows the child's overall accuracy, their accuracy on type of questions, and the proportion of each question type in this session (F17).

In addition, the parent can check the child's weekly progress. As shown in Figure 8, the dashboard allows them to review the child's weekly progress and the overall accuracy in question-answering (F16). They can also tailor the dashboard to display the statistics of a particular question type. Besides, the dashboard also informs the parent of the proportion of each question type from all sessions in the week (F16).

5.5 Implementation

5.5.1 Web Application. The front-end interactive web application of StoryBuddy is implemented in React and hosted using Python's built-in HTTP server. The web-based nature of StoryBuddy allows it to run from the web browsers on a variety of devices including desktops, laptops, tablets, and smartphones. The use of React allows it to be "responsive" so that its graphical user interfaces can adjust to fit different screen dimensions and ratios. For all functionalities to work properly, StoryBuddy requires the device to have a microphone and a speaker.

StoryBuddy uses Google's Cloud Text-to-Speech API[1] for speech synthesis in story-reading, which yields natural-sounding results. Storybooks in StoryBuddy are stored in a simple JSON format, which allows users and community members to easily add new storybooks to its story library.

5.5.2 Conversational Agent. The conversational agent used in facilitating agent-child reading is implemented with the *react-simple-chatbot* framework[2] at the front end for displaying the chat history and facilitating input/output. At the back-end, it uses the Google Dialogflow[3] framework for intent detection and classifying the child's answers to determine their correctness. We trained the answer classification model in Dialogflow with a small rule-based corpus. Given a model-generated answer (Section 5.5.3), we proliferated the Dialogflow training phrases of each question by applying templates upon its answer, so that the agent can correctly handle variations in the answer (e.g., "three bears" vs. "3 bears") as well as fillers in the answer such as "*It may be <answer>*", "*I believe <answer>*", and "*I guess <answer>*". Note that the Dialogflow training only takes a few seconds, therefore when a new question is generated, the chatbot becomes ready to respond to the user's answer in real-time.

5.5.3 Question Generation Model. For question generation, StoryBuddy uses an automated question-answer generation (QAG) model trained on the FairytaleQA dataset [87]. This QAG model can automatically generate high-quality QA pairs from any children's storybooks. The questions generated by the QAG model are designed to mimic the style "as if a teacher or parent is to think of a question to improve children's language comprehension ability

---

[1]https://cloud.google.com/text-to-speech
[2]https://github.com/LucasBassetti/react-simple-chatbot
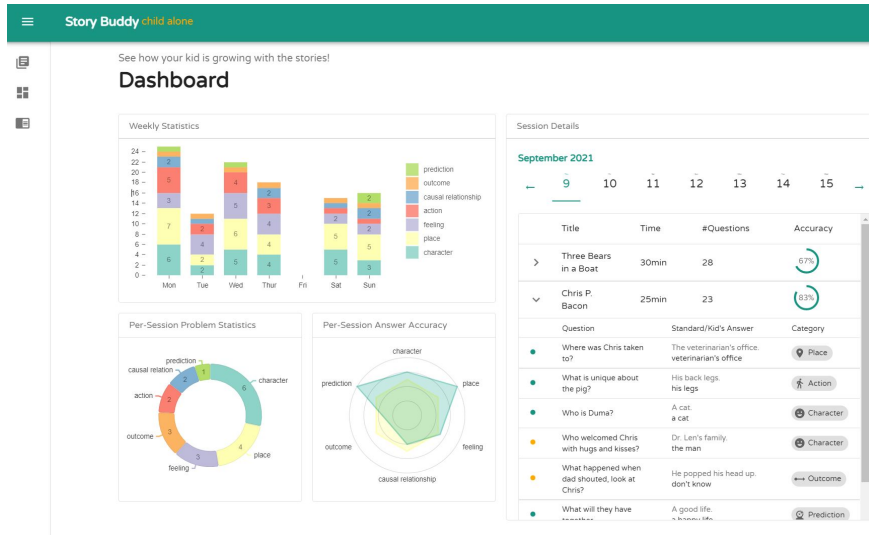[3]https://cloud.google.com/dialogflow

Figure 8: StoryBuddy's dashboard for child progress tracking and performance assessment

while reading a story to them [87]". This QAG model also supports generating questions and the corresponding answers from a specific question type (detail of the supported types in Section 5.3).

On a high level, the pipeline of this QAG model consists of a rule-based answer generation module, a BART-based question generation module [37], and a ranking module. The QAG model was trained on the FaiytaleQA dataset, which contains 922 QA pairs from 46 children's storybooks labeled by education experts. In a human evaluation, this QAG model achieved state-of-art performance in generating high-quality question-answer pairs from children's storybooks [87].

In order to identify follow-up questions, StoryBuddy groups a large amount of QAG-generated questions for each section. The strategy that StoryBuddy uses is to treat the top-3 questions as anchors and then calculate the similarity between each of the remaining questions and the anchored questions. We define similarity as the number of overlapping tokens after removing the stop words. A question will become a candidate for the follow-up question of an anchored question if the similarity between them is greater than 3 and the answer of the question is not included in the anchored question text. An anchored question will not have a follow-up if it lacks an eligible candidate.

6 USER STUDY

We conducted a remote user study[4] to evaluate StoryBuddy. The study examined the following research questions:

- **RQ1**: Can parents successfully use StoryBuddy to create interactive storytelling experiences for their children?
- **RQ2**: How do parents and children interact with StoryBuddy in its two modes?
- **RQ3**: Do parents and children find StoryBuddy usable, useful, and likable?

---
[4]The study protocol was approved by the IRB at our institution.

6.1 Participants

We recruited 12 pairs of participants (PB1–PB12) from university mailing lists and through the snowball sampling method [53]. Each pair consisted of a parent and a child between the ages of 3–8. All the participants resided in the U.S. and were fluent in English. Eight parents used English as their second language. The demographic characteristics of the participants are reported in Table 2. Each pair of participants was compensated with a $50 gift card for their time.

6.2 Study Procedure

Each user study session lasted around an hour and was conducted remotely over Zoom due to the impact of the COVID-19 global pandemic. Participants accessed StoryBuddy using the browser on their own computers and shared their screens with the experimenter. Participants were also encouraged to turn on their cameras if possible. All user study sessions are video recorded.

Prior to the beginning of each session, the parent signed the consent form and filled out a demographic questionnaire. After the experimenter gave a short introduction to the study, the participants watched a 4-minute tutorial video on how to use StoryBuddy.

Each pair of participants then used StoryBuddy to read two stories: *Three Little Bears* and *Chris P. Bacon: My Life So Far* in two modes (*parent-AI co-reading* and *automated bot-reading*). The two stories were chosen because of their appropriate lengths for the study and appropriate difficulties for the target age group. The order of the stories, as well as the match between the modes and the stories, were random. In the parent-AI co-reading mode, the parent read the story to their child using the story reading interface and facilitated interactive question-answering with the help of StoryBuddy, as described in Section 5.2.1. In the automated bot-reading mode, the parent first customized the questions to be used by the agent using the configuration page. They then used the StoryBuddy agent to automatically read the story and interact with the child with questions, as described in Section 5.2.2. We asked the

ID	Parent's gender	Child's gender	Parent's age	Child's age	Storytelling frequency	ESL?
PB1	Female	Female	45-54	5	More than 6 times a week	N
PB2	Female	Female	35-44	7	More than 6 times a week	N
PB3	Male	Male	25-34	5	More than 6 times a week	Y
PB4	Female	Female	25-34	4	1-3 times a week	N
PB5	Female	Female	35-44	7	4-6 times a week	Y
PB6	Female	Male	25-34	4	More than 6 times a week	Y
PB7	Female	Female	25-34	7	More than 6 times a week	Y
PB8	Female	Male	25-34	5	1-3 times a week	Y
PB9	Male	Male	25-34	5	1-3 times a week	Y
PB10	Female	Female	35-44	7	Not reported	N
PB11	Female	Female	35-44	6	More than 6 times a week	Y
PB12	Female	Female	35-44	6	4-6 times a week	Y

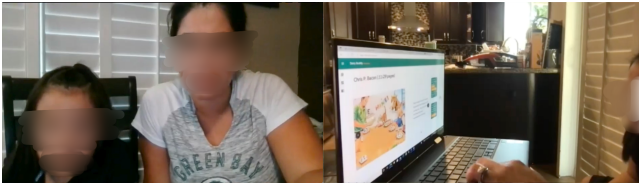**Table 2: Demographics of user study participants.**

Figure 9: A screenshot from a remote user study session, showing a child and her parent interacting with StoryBuddy

parent to avoid intervening when the child was interacting with the agent in the automated bot-reading mode. Figure 9 illustrated the scenario of how the parent and the child interacted StoryBuddy in our study: the parent sat beside the child and they used StoryBuddy to read the assigned story in one of the modes. They could chat with each other during the study.

After trying out StoryBuddy, we conducted a 10-minute semi-structured interview with each participant on their experience interacting with StoryBuddy.

6.3 Study Results

All 12 pairs of participants successfully completed the assigned parent-AI co-reading and automated bot-reading sessions. The parent-AI co-reading session and the automated bot-reading session lasted around 18 and 17 minutes on average respectively.

6.3.1 Observation of user behaviors. To understand how parents and children used the StoryBuddy in the study, we analyzed the observed user behaviors from screen recordings in the study. We aim to answer the following questions: (1) How did parents read story with kids in the parent-AI co-reading mode? (2) How did parents ask questions in the parent-AI co-reading mode? (3) How did parents use follow-up questions in the parent-AI co-reading mode? (4) How did parents configure the questions in the child-alone mode? (5) How did children interact with the chatbot in the child-alone mode?

Analysis methods. Based on the above research questions, we came up a list of behaviors of interest for the annotation (Table 3). Note that those behaviors are rather objective (e.g. the user interacts

with a certain feature in the system) with little space for subjective interpretation. One author carefully went through the screen recording videos of all study sessions to annotate these behaviors and count their occurrences.

Findings. Parents used different reading strategies in the parent-AI co-reading mode. A factor that affected the reading strategy of choice was the age of their children. Five parents used the auto-reading feature throughout the study. Four parents who all have younger children (5 or younger) decided to read the story by themselves. Compared with the text-to-speech auto-reading, they deliberately made the reading more emotional and slower. On the contrary, three parents of the older children from 6–7 had their children lead the story reading and only helped them with new words. These findings confirmed the usefulness of supporting multiple reading strategies (and a mix of them) in Storybuddy. Such differences may be attributed to the differences between children in the pre-reading stage and the early-reading stage, where younger children rely on the sound structure of spoken language but older children can read stories more independently [55]. It also indicates the opportunity for supporting better age-based customization, which we will discuss in Section 8.

Parents used the generated questions in different ways. Two parents asked all the displayed questions in the given order on every page. Seven parents only selected some questions to ask, when we ask them about how they picked the questions to use in the post-study interview, they reported that their selections were based on their intuition of whether the questions were comprehensible to the kids. For the generated follow-up questions, five parents did not ask them at all throughout the reading. Four parents asked those questions sometimes. They tended to skip follow-up questions when they noticed the child became impatient, or when those questions were less relevant with previous ones. Lastly, two parents used the chatbot to ask follow-up questions.

For the question type configuration in the child-alone mode, six parents directly used the default settings without any modification. The other six parents reviewed and made some modifications to the preferred question types to be generated. This implies that some parents may not want to bother with the configuration, while

Research questions	Behaviors of interest
How did parents read story with kids in the parent-AI co-reading mode?	The parent uses the auto-reading feature
	The parent reads the story by themselves
	The parent lets the child read the story
How did parents ask questions in the the parent-AI co-reading mode?	The parent asks a generated question themselves
	The parent uses the chatbot to ask a generated question
	The parent asks a question of their own
How did parents use follow-up questions in the parent-AI co-reading mode?	The parent asks a provided follow-up question by themselves
	The parent uses the chatbot to ask a follow-up question
	The parent asks a follow-up question of their own
How did parents configure the questions in the child-alone mode?	The parent uses the default setting
	The parent configures the preferred types of generated questions
How did children interact with the chatbot in the child-alone mode?	The child answers an initial question from the bot
	The child reattempts a question when the answer was incorrect
	The child answers a follow-up question from the bot

Table 3: The behaviors of interest from screen recordings

others take advantage of the configuration option to better adapt the system to their goals and needs.

We also analyzed how children interacted with the chatbot in the child-alone mode. There was one child who simply used the Story-Buddy to read the book and skipped all interactions with the chatbot. All other eleven children interacted with the chatbot. However, the extent to which they used the chatbot varied. Six of them only tried to answer the first question throughout the study and would move on to the next page no matter whether their answer was considered correct by the chatbot. Among the rest, three children were willing to make another attempt when their answers were incorrect, but would skip to the next page if their reattempt was still not accepted. Only two children kept interacting with the chatbot on a question until the answer got accepted. These observations indicate future design opportunities in more effectively engaging with children in multi-turn conversations to provide them with helpful guidance (especially in case of partially correct answers) that assists them in refining their answers.

6.3.2 Post-study interview. We ended each study session with a 10-minute semi-structured interview with the parent. Besides following up on their post-study questionnaire responses, we further asked the parent about the difficulties they encountered during the study, whether and how the StoryBuddy helped the children and parents on storytelling, how they would use the StoryBuddy in their daily life, and their suggestions on system improvement.

Analysis methods. In the guidance of established open coding methods [10, 34], two authors conducted a thematic analysis of interview transcripts to common themes with respect to user experiences, challenges, potential usage, and feedback. Specifically, each coder first individually went through and coded the transcripts of all the interview sessions using an inductive approach. For user quotes that did not include straightforward key terms, coders assigned researcher-denoted concepts as the code. Two coders discussed the themes emerging from the coding process and reached a consensus on the codebook. Then they independently mapped the extracted

quotes to codes. The two coders reached a strong level of agreement in the interrater reliability with Cohen's Kappa $\kappa = 0.81$.

StoryBuddy as a useful language-learning tool for non-English speaking families. Four parents who were not native English speakers recognized the use of StoryBuddy for facilitating language learning through AI-facilitated story reading in non-English speaking families. The StoryBuddy system can teach children the "*correct pronunciation*" (PB11) and "*vocabulary*" (PB8) that parents were not good at. Besides, the StoryBuddy system allowed young children to enjoy storytelling even if the caregivers cannot speak English. For example, PB5 said "*I'm also thinking because we're bilingual family, and some of our family members, like grandparents, they cannot speak English. But if grandparents are caregivers, and they use this system, they can also still help their grandchild to do the English story reading.*" These findings confirmed results from prior study on how conversational agents can support children's language development by serving as their language partners [81].

Using StoryBuddy to develop and assess children's reading skills. Seven parents commented on StoryBuddy's value in developing and assessing children's reading skills. Specifically, they thought the generated questions for stories can "*cultivate children's critical reading*" (PB1) and help them "*figure out what the key points were of the paragraph they were reading*" (PB2). Plus, one participant (PB4) believed StoryBuddy can engage his kid and make him more concentrated during the story-reading: "*if there's actually something like this that's able to help them (the children) be more engaging, they will be more concentrated on the story they are reading*" (PB4). On the assessment side, the dashboard provides parents with "*a straightforward way to understand the children's current performance*" (PB6) and "*know what aspects of reading skills their children need help on*" (PB8). PB9 said, "*the dashboard is pretty good. It's kind of capture what does the kid like or dislike (about question and story), and what they learn.*" Such objectives of StoryBuddy reported by participants are consistent with prior literatures on how conversational agents

can serve as effective partners that support the improvement of children's story comprehension [82].

Reducing parents' burdens in storytelling. Ten parents confirmed that StoryBuddy can reduce their burden by allowing flexible involvement and assisting them in coming up with questions, which was in the initial design goals (DS2 and DS6) and came from an original insight in the participatory design (Section 4.2.4). Story-Buddy was found to be especially useful in time-conflict situations when "*they (children) want to read something but you (parents) are not available*" (PB10), or be "*a supplementary tool when the parents feel tired or lazy*" (PB3). However, PB3 also mentioned that the StoryBuddy cannot totally replace the parents' role in storytelling, because "*having a reading time together is a special time of enhancing parent-child relationship and those emotional parts cannot be offered by machine.*" Nonetheless, parents agreed that the Story-Buddy would "*save their effort on coming up with questions*" (PB5) and this embedded question-answering feature could "*make the storytelling more engaging*" (PB4).

Personalizing the reading experience. Nine parents commented on StoryBuddy's features for personalization, especially through its support of parent question configuration. They liked that they were able to control the types of generated questions, review generated questions and made changes as needed. Parents also pointed out several opportunities for further enhancing StoryBuddy's support of personalization. For example, PB10 advised that StoryBuddy can serve as a book recommender: "*since not every parent and children know the proper (difficulty) level, you can make this system recommend the books to kids, perhaps also based on their interests and reading history*", though it may require the StoryBuddy to "*collect more book selections covering a wide range of topics*" (PB10). Parents also suggested that the interaction flow of StoryBuddy could be more flexible to fit different family's story-reading practices. For example, PB3 preferred to "*have the question-answering session at the end of the whole book instead of on every page*" (PB3) and the current design could "*break the natural flow of parent-child storytelling*" (PB3). While another parent really liked the questions-at-every-page approach: "*I liked the way that the questions being asked because it happened on every page. There are actually similar applications that read the whole story and ask questions at the end. In this way, the kids could easily forget the content and it is harder for them to answer the question (Her kid was nodding). I think this (the in-place question-asking) is the best feature you did in your app*". These findings confirmed the importance of enabling the system to adapt to the unique contexts, preferences, and needs of parents and children as well as the established family practice in AI-enabled systems that facilitate parent-child interactions.

Opportunities for adapting StoryBuddy to different age groups. The interviews revealed opportunities for further personalizing the reading experience offered by StoryBuddy based on the age and the developmental stage of children. Although the dialogic reading approach is generally effective for children aged three to eight, which is the age group that StoryBuddy targets, it would be useful to further customize the reading experience for children of different ages within this group. For some older children, some generated questions could be "*too stupid and meaningless so that the kids lost their interest in using this system*" (PB7), while the younger ones found a small number of questions were "*very hard and beyond understanding*" (PB8). The age-based personalization can also apply to interaction strategies. The bot-guided reading approach in the parent absent mode generally worked quite well for younger children in our study, but the guidance may be "too slow" or even annoying for older children or children who are advanced in the development of reading skills, as PB12 said "*I think it may be useful for those younger kids, my son is able to read stories on his own, he may not need such a system to assist his reading.*" Such differences correspond to the distinctions between children in the pre-reading stage and those in the early-reading stage, where older children start to learn how to read independently without relying on the sound structures of spoken language [25, 55].

Usability issues. The general attitude towards the usability of Sto-ryBuddy was positive. However, the participants still experienced some specific usability issues during the use. The first problem was the inaccurate speech recognition in the Google Cloud Speech Recognition API. It sometimes misinterpreted the children's speech or stopped the recognition early. As PB9 said "*the children would not be frustrated because of the incorrect recognition of their answers*". Parents also found the size of some widgets in StoryBuddy were too small for their kids: "*the buttons were too small for children to use, especially when they use fingers on iPad*" (PB9). Besides, parents also suggested to "*make the interface more colorful or add more cartoon element*" (PB3) so as to engage kids to use.

6.4 Threats to Validity

A potential threat to the validity of the user study results is the sample bias in our participants. Since our participant recruitment was done in a University community, all the participating parents had at least a bachelor's degree. More than half of the participating parents had received or were pursuing a graduate degree. Parents in underrepresented racial groups, from lower socioeconomic backgrounds, or with non-traditional family structures were also underrepresented in our participants.

Another threat lies at the ecological validity—The study was done remotely through video calls and screen-sharing via Zoom. The setting of the study did not closely resemble the realistic setting where StoryBuddy might be used. For example, when we tested the automated bot-reading mode of StoryBuddy, the parent was asked to refrain from intervening the child-agent interaction. However, the parent was still co-located and within the sight of the child. In realistic usage of this mode, the parent will likely be completely absent.

It would be also interesting to investigate how much of a role did the *novelty factor* of interacting with a conversational agent played in the strong child engagement during the storytelling sessions in our user study and to measure how the effect of such novelty factor would change over time.

We plan to address these potential threats to validity in the future through (1) first, a larger-scale field deployment study with a more representative user population; and (2) eventually, a public release of StoryBuddy to the general public (detail in Section 8).

7 DISCUSSION AND DESIGN IMPLICATIONS

The results from our user study suggest that parents can successfully collaborate with StoryBuddy to create interactive storytelling experiences for their children. StoryBuddy also performed well in keeping the child engaged and entertained. Its interaction design allowed changes in how the child interacts with the parent and the agent throughout the storytelling process so that the child did not easily get bored.

Below, we discuss the implications and design themes emerged from our work [3, 26].

7.1 Partial Automation in Parent-AI Collaboration

From the lens of mixed-initiative interaction [27] and human-AI collaboration [75, 76], a central issue to consider in designing parent-AI collaborative systems is to identify opportunities for *partial automation* based on the capabilities of the AI system and the capabilities of the parent [33]. For example, in StoryBuddy, the key AI capability we leveraged is that its back-end model can quickly generate questions, answers, and possible follow-up questions of seven types from the textual content of *any* story. However, it is important to recognize that the parent possesses knowledge about the task and the context that is unknown to the system. For instance, they know the preferences of their child, their own goals for child skill development, and the constraints in the current context. Such knowledge allows the parent to customize the generated questions accordingly.

To accommodate partial automation, the interaction flow of the system needs to cope with interruptions and resumptions. In the opposite direction, the parent also needs to be attentive and ready to contribute when needed. For example, in the parent-AI co-reading mode of StoryBuddy, the parent needs to make on-the-fly decisions on whether to use an AI-recommended question or an AI-generated follow-up question in synchronous storytelling. However, since the parent is also the one telling the story in this mode, they are able to adjust the pace of the parent-child interaction. If we had designed a system where the system took initiative in telling the story but the parent needed to make decisions on what questions to ask and how to ask them while keeping up with the pace of the system, coordinating harmonious collaboration between the AI system and the parent would be much more challenging.

7.2 Supporting Flexibility in Parent Involvement

Beyond accounting for the capabilities of the parent and the AI system when designing partial automation, we also need to consider the parent's varied preferences and changing availability in their involvement. For example, in our formative study, PD process, and user study, we encountered parents who cared deeply about the educational goals and skill assessment features in StoryBuddy. They wished to have fine granularity control on the types of questions generated by the system and the specific follow-up questions to use. For them, we designed the question preference configuration panel (Section 5.3), the interactive dashboard for child progress tracking and performance assessment (Section 5.4), and the features for manually editing questions and answers when configuring the StoryBuddy agent in the automated bot-reading mode. However,

the use of all these features is optional—a parent who mainly uses StoryBuddy for engagement and entertainment goals could skip all these steps if they wish so.

Another design consideration for parent involvement is how we can help parents stay involved when they are unavailable for synchronous parent-child interactions. The conflict between (1) parents' desire for fulfilling their children's storytelling needs and staying involved in the process; and (2) parents' limited and constrained time availability was a recurring theme in our studies. This problem is aggravated by the increasingly common work-from-home arrangement for parents. In work-from-home situations, parents may *seem available* to children for interaction since they are physically home when they are, in fact, unavailable. An interactive storytelling agent that can keep children engaged without requiring parental intervention would be particularly useful in such scenarios. To address this issue, the core strategy that StoryBuddy used is to turn synchronous involvement into asynchronous involvement— While the delivery of the story content and the interaction for question-answering is facilitated by the StoryBuddy agent, the parent has (1) control over the content of the interaction through the configuration *before* a storytelling session; and (2) knowledge on the progress and the performance of the child through viewing the dashboard *after* a storytelling session.

7.3 Role of AI in Parent-Child Interaction

The design of StoryBuddy introduced an AI system into an *existing* activity that previously involved only the parent and the child: parent-child joint reading, as a result, an important challenge is to determine the appropriate role that the AI should play. Should the AI system be an assistant, a peer, a companion, an agent of the parent, or something else? To complicate the problem, the activity of parent-child joint reading usually fulfills a combination of multiple goals: relationship building, skill development and assessment, and entertainment. Therefore, it is crucial to consider the roles that the AI system would play in each part of the activity and how its roles contribute (positively or negatively) to the user's goals.

For example, in our design, StoryBuddy assists the parent by helping them identify opportunities for asking questions, recommending questions to use, and proposing follow-up questions. These forms of assistance help parents come up with better questions for fulfilling their skill development and assessment goal, reduce their cognitive load so that they can allocate more attention to the interaction with their child for the relationship-building goal, and keep their child engaged and entertained through the occasional interaction between the child and the agent. However, all these forms of AI involvement do not lessen the central role of the parent in their interactions with the child. In its chatbot form, StoryBuddy starts to act as a companion or a peer to children as it communicates with them through natural language dialogs—the agent may also be leveraged as a third-party mediator that facilitates parent-child communication as reported in [4].

There are still issues with the current design of StoryBuddy and opportunities for how its roles can be refined. For example, as discussed in Section 6.3.2, a parent criticized that StoryBuddy interfered with their normal story reading flow by proposing to ask questions after every section instead of at the end of the whole story

(which they often do in their parent-child joint reading practice without AI involvement). Parents also suggested that StoryBuddy may further help with the entertainment and engagement goal by combining its tracking of child progress with gamification. For example, it can give out virtual "badges" when the child reaches certain achievements and milestones.

In future projects on designing AI systems for parents and children, it is crucial to first start with formative studies that uncover the multi-faceted goals of the parent and the child. After that, multiple design iterations with intensive user participation in the user-centered design process are needed in order to carefully define the appropriate roles that the AI system should play.

8 FUTURE WORK

There are several directions for planned future work. The current version of StoryBuddy focuses on parent-led and agent-led question-answering. A different variation of question-answering that we plan to support in the future is child-led question-answering, where the child asks questions about the story plot and the agent can answer these questions and ask appropriate follow-up questions. A parent in our PD session reported currently using this strategy in their parent-child joint story-reading (without the use of digital assistance) with success. Prior literature [72] suggested that a bot that can answer children's questions during the storytelling process is preferred by children. We may also explore the design space of other interaction approaches in interactive storytelling, such as reflective storytelling [28]. Another useful enhancement on StoryBuddy's question generation capabilities is to support image-based questions about the content of the visual illustrations in the storybook in addition to questions about the textual contents. This idea came up in our discussion with multiple parents in both the PD process and the usability study.

While the current version of StoryBuddy supports a wide range of device types including smartphones, tablets, laptops, and desktops, another type of device that would be useful to support is the smart speaker. Some parents expressed their concerns about limiting the "screen time" for their children in our studies. The child may also, for example, switch away from StoryBuddy and play video games instead when they use StoryBuddy on a smartphone or a tablet when the parent is absent. Supporting running StoryBuddy on smart speakers such as Amazon Echo, Apple HomePod, or Google Home can alleviate these concerns. However, adapting StoryBuddy to a screenless smart speaker raises additional design challenges in, for example, maintaining conversational context in longer multi-turn conversations, facilitating turn-taking in question-answering, grounding questions back to the story content, and displaying the progress of storytelling sessions. We plan to investigate these issues for future work.

A direction of future work informed by participatory design and user study results is to make the back-end question generation model adaptive to parent preferences and child interactions. With the current version of StoryBuddy, the parent can customize the questions through (1) configuring the back-end question generation model on the preferred question types (Section 5.3); and (2) selecting and editing the generated questions for each story (Figure 5). However, this process can get tedious if the parent wishes

to go through all the questions. Once the configuration is done, the agent's question-asking plan also remains static without the ability to adjust based on child interactions. To address these limitations, we plan to explore inserting a model in StoryBuddy that learns the parent's preferences while they select and tweak the questions to use. After the parent finishes configuring a few pages, the model can automatically adjust the generating questions for the rest of the story according to the learned parent preferences to reduce parent effort. At the story-telling time, another model can track child performance and engagement in order to automatically adjust the difficulty and the types of questions in real-time.

Another direction for future study is to understand how children from different age groups interact with StoryBuddy differently and how the design of StoryBuddy can be extended to accommodate their diverse needs. As reported in the results from the post-study interviews in Section 6.3.2, parents identified several opportunities in the current version of StoryBuddy to better accommodate the needs of children of different ages: on the question-generation back-end, the model could adjust the complexity, the vocabulary used, and the cognitive skill required in its generated questions based on the development progress of the child. The interactive strategy on e.g., how much guidance to provide during story reading in the parent absent mode can also vary according to the needs of children of different ages. Prior literature also suggested that children from the pre-reading and early-reading groups enjoy and benefit from graphical contents (e.g., pictures) in storybooks as much as textual contents [13, 54]. While the current version of StoryBuddy displays pictures from the original book in its story reading interface, none of its interactions with parents and children refers to the content of the pictures. For future work, it would be useful to investigate ways to generate questions about pictures in the storybook or questions that connect the textual contents of a storybook with its graphical contents.

In addition, our user study focused on using StoryBuddy in the home settings given the important role home literacy environments play in children's literacy development. We are also interested in exploring how automatic question generation can be used to support teachers in classroom instructions. It is conceivable that our system has the potential to enable more personalized interactive reading instruction. First, teachers can use our system to easily generate customized reading resources for students based on student needs. Second, teachers can better monitor students' reading comprehension by tracking the students' performance in answering dialogic questions in real-time. This will allow the early intervention of students who may be at risk for reading difficulty.

Another currently under-explored design space in StoryBuddy is the child-led intervention. The current support interactions in StoryBuddy are mostly driven by parents—parents prepare the co-reading companion by configuring the question types and selecting the mode to use. In the parent-child co-reading mode, the parent is the one who controls the flow of the co-reading and decides when to ask a question and what question to ask. The child-led approach, as described in [68], has the potential to directly engage with children and empower children to learn by themselves rather than relying on adults. In order to support this in future versions of StoryBuddy, we need to improve the design of configuration interfaces with the capability and the preferences of children in

mind so that they can configure the StoryBuddy agent themselves. Children may also use assistance in e.g., recommending a potential storybook of interest. Future design activities are also needed to explore how to foster relationships and rapport-building between children and the AI-enabled story-reading companion in a child-led rather than parent-driven fashion. To take another step forward, another direction to further support child-led interaction is to involve children in the creation and development of stories. Such collective storytelling among children, AI systems, and (optionally) parents can give children structure and significance to the world around them, facilitating pedagogical and psychological development [3]

Lastly, we are hoping to deploy StoryBuddy with a larger group of users and eventually release the system for public use. Although the design of StoryBuddy was informed by formative study and participatory design results, and the usability of StoryBuddy has been validated by a user study, we hope to further understand how parents configure different kinds of stories for children, how parents choose between the two modes, how parents use the different interaction paradigms in the parent-AI co-reading mode, and how useful StoryBuddy is for users in realistic contexts. The main goal of the deployment would be to study StoryBuddy within its intended context of use. Another goal is to study the use of StoryBuddy in a more representative user population. As discussed in Section 6.4, there are demographic biases in the participant population of our user study. We seek to recruit a more diverse group of users in the deployment and public release of StoryBuddy.

9 CONCLUSION

This paper presented StoryBuddy, a new system that allowed parents to collaborate with an AI system in creating storytelling experiences with interactive questioning-answering. Informed by results from a formative study and a participatory design study, we designed two distinct modes and several dynamic interaction paradigms that supported flexible degrees of parent involvement in (1) the configuration of the interactive storytelling experience before the storytelling; (2) the parent-AI co-delivery of the story content to the child during the storytelling; and (3) the tracking and assessment of child progress and performance after the storytelling. A user study with 12 pairs of parents and children found that StoryBuddy was effective in providing parents with desired levels of control and involvement while maintaining children's engagement in the storytelling process. Parents and children considered StoryBuddy useful, helpful, and likable.

REFERENCES

[1] Diogenis Alexandrakis, Konstantinos Chorianopoulos, and Nikolaos Tselios. 2019. Insights on Older Adults' Attitudes and Behavior Through the Participatory Design of an Online Storytelling Platform. In *Human-Computer Interaction – INTERACT 2019*, David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris (Eds.). Springer International Publishing, Cham, 465–474.

[2] Zahra Ashktorab, Mohit Jain, Q Vera Liao, and Justin D Weisz. 2019. Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 254.

[3] Wolmet Barendregt, Olof Torgersson, Eva Eriksson, and Peter Börjesson. 2017. Intermediate-Level Knowledge in Child-Computer Interaction: A Call for Action. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) *(IDC '17)*. Association for Computing Machinery, New York, NY, USA, 7–16. https://doi.org/10.1145/3078072.3079719

[4] Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2020. *Parenting with Alexa: Exploring the Introduction of Smart Speakers on Family Dynamics*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376344

[5] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 91 (Sept. 2018), 24 pages. https://doi.org/10.1145/3264901

[6] Neelma Bhatti, Timothy L Stelter, and D Scott McCrickard. 2021. Conversational User Interfaces As Assistive interlocutors For Young Children's Bilingual Language Acquisition. *arXiv preprint arXiv:2103.09228* (2021).

[7] Melanie Birks, Ysanne Chapman, and Karen Francis. 2008. Memoing in qualitative research: Probing data and processes. *Journal of research in nursing* 13, 1 (2008), 68–75.

[8] Pamela Blewitt, Keiran M Rump, Stephanie E Shealy, and Samantha A Cook. 2009. Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology* 101, 2 (2009), 294.

[9] Susanne Bodker. 1999. Scenarios in user-centred design-setting the stage for reflection and action. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. IEEE, 11–pp.

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3 (2006), 101 – 77.

[11] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. ACM, New York, NY, USA, Article 43, 12 pages. https://doi.org/10.1145/3098279.3098539

[12] Griffin Dietz, Jimmy K Le, Nadin Tamer, Jenny Han, Hyowon Gweon, Elizabeth L Murnane, and James A. Landay. 2021. *StoryCoder: Teaching Computational Thinking Concepts Through Storytelling in a Voice-Guided App for Children*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445039

[13] DR Donald. 1983. The use and value of illustrations as contextual information for readers at different progress and developmental levels. *British Journal of Educational Psychology* 53, 2 (1983), 175–185.

[14] Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. Unified language model pre-training for natural language understanding and generation. *arXiv preprint arXiv:1905.03197* (2019).

[15] Xinya Du, Junru Shao, and Claire Cardie. 2017. Learning to Ask: Neural Question Generation for Reading Comprehension. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. ACL, Vancouver, Canada, 1342–1352. https://doi.org/10.18653/v1/P17-1123

[16] Yao Du, Kerri Zhang, Sruthi Ramabadran, and Yusa Liu. 2021. *"Alexa, What is That Sound?" A Video Analysis of Child-Agent Communication From Two Amazon Alexa Games*. Association for Computing Machinery, New York, NY, USA, 513–520. https://doi.org/10.1145/3459990.3465195

[17] Zoe M Flack, Andy P Field, and Jessica S Horst. 2018. The effects of shared storybook reading on word learning: A meta-analysis. *Developmental psychology* 54, 7 (2018), 1334.

[18] David M. Frohlich, Dorothy Rachovides, Kiriaki Riga, Ramnath Bhat, Maxine Frank, Eran Edirisinghe, Dhammike Wickramanayaka, Matt Jones, and Will Harwood. 2009. *StoryBank: Mobile Digital Storytelling in a Development Context.* Association for Computing Machinery, New York, NY, USA, 1761–1770. https://doi.org/10.1145/1518701.1518972

[19] Neil Frude and Steve Killick. 2011. Family storytelling and the attachment relationship. *Psychodynamic Practice* 17, 4 (2011), 441–455.

[20] Radhika Garg and Subhasree Sengupta. 2020. Conversational Technologies for In-home Learning: Using Co-Design to Understand Children's and Parents' Perspectives. In *Proceedings of the 2020 CHI conference on human factors in computing systems.* 1–13.

[21] Radhika Garg and Subhasree Sengupta. 2020. *Conversational Technologies for In-Home Learning: Using Co-Design to Understand Children's and Parents' Perspectives.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376631

[22] Jonathan Grudin and John Pruitt. 2002. Personas, Participatory Design and Product Development: An Infrastructure for Engagement. In *Proceedings of Participation and Design Conference (PDC2002), Sweden.* 144–161. http://www.itee.uq.edu.au/~comp4501/_2003/_Readings/GrudinPersonas.pdf

[23] Anne C Hargrave and Monique Sénéchal. 2000. A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly* 15, 1 (2000), 75–90.

[24] Felix Hill, Antoine Bordes, Sumit Chopra, and Jason Weston. 2015. The goldilocks principle: Reading children's books with explicit memory representations. *arXiv preprint arXiv:1511.02301* (2015).

[25] Torleiv Hoien and Ingvar Lundberg. 1988. Stages of word recognition in early reading development. *Scandinavian Journal of Educational Research* 32, 4 (1988), 163–182.

[26] Kristina Höök and Jonas Löwgren. 2012. Strong Concepts: Intermediate-Level Knowledge in Interaction Design Research. *ACM Trans. Comput.-Hum. Interact.* 19, 3, Article 23 (oct 2012), 18 pages. https://doi.org/10.1145/2362364.2362371

[27] Eric Horvitz. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99).* ACM, New York, NY, USA, 159–166. https://doi.org/10.1145/302979.303030

[28] Layne Jackson Hubbard, Yifan Chen, Eliana Colunga, Pilyoung Kim, and Tom Yeh. 2021. *Child-Robot Interaction to Integrate Reflective Storytelling Into Creative Play.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3450741.3465254

[29] Young-Suk Grace Kim. 2017. Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading* 21, 4 (2017), 310–333.

[30] Jacqueline Kory and Cynthia Breazeal. 2014. Storytelling with robots: Learning companions for preschool children's language development. In *The 23rd IEEE international symposium on robot and human interactive communication.* IEEE, 643–648.

[31] Huseyin Kotaman. 2020. Impacts of dialogical storybook reading on young children's reading attitudes and vocabulary development. *Reading Improvement* 57, 1 (2020), 40–45.

[32] Igor Labutov, Sumit Basu, and Lucy Vanderwende. 2015. Deep Questions without Deep Understanding. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers).* ACL, Beijing, China, 889–898. https://doi.org/10.3115/v1/P15-1086

[33] Tessa Lau. 2009. Why Programming-By-Demonstration Systems Fail: Lessons Learned for Usable AI. *AI Magazine* 30, 4 (Oct. 2009), 65–67. http://www.aaai.org/ojs/index.php/aimagazine/article/view/2262

[34] Jonathan Lazar, Jinjuan Feng, and Harry Hochheiser. 2010. Research Methods in Human-Computer Interaction.

[35] Pillar Learning. 2021. Meet Codi - An Interactive, AI-Enabled Smart Toy for Kids! https://www.pillarlearning.com/

[36] Rosemary Lever and Monique Sénéchal. 2011. Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of experimental child psychology* 108, 1 (2011), 1–24.

[37] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461* (2019).

[38] Toby Jia-Jun Li, Amos Azaria, and Brad A. Myers. 2017. SUGILITE: Creating Multimodal Smartphone Automation by Demonstration. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17).* ACM, New York, NY, USA, 6038–6049. https://doi.org/10.1145/3025453.3025483

[39] Toby Jia-Jun Li, Jingya Chen, Haijun Xia, Tom M. Mitchell, and Brad A. Myers. 2020. Multi-Modal Repairs of Conversational Breakdowns in Task-Oriented Dialogs. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST 2020).* ACM. https://doi.org/10.1145/3379337.3415820

[40] Toby Jia-Jun Li, Tom Mitchell, and Brad Myers. 2020. Interactive Task Learning from GUI-Grounded Natural Language Instructions and Demonstrations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations.* ACL, Online, 215–223. https://doi.org/10.18653/v1/2020.acl-demos.25

[41] Chaolan Lin, Selma Šabanović, Lynn Dombrowski, Andrew D. Miller, Erin Brady, and Karl F. MacDorman. 2021. Parental Acceptance of Children's Storytelling Robots: A Projection of the Uncanny Valley of AI. *Frontiers in Robotics and AI* 8 (2021), 49. https://doi.org/10.3389/frobt.2021.579993

[42] Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out.* ACL, Barcelona, Spain, 74–81. https://aclanthology.org/W04-1013

[43] Ling. 2021. Luka AI reading companion. https://luka.ling.ai/

[44] Silvia B Lovato, Anne Marie Piper, and Ellen A Wartella. 2019. Hey Google, do unicorns exist? Conversational agents as a path to answers to children's questions. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children.* 301–313.

[45] Silvia B. Lovato, Anne Marie Piper, and Ellen A. Wartella. 2019. Hey Google, Do Unicorns Exist? Conversational Agents as a Path to Answers to Children's Questions. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19).* Association for Computing Machinery, New York, NY, USA, 301–313. https://doi.org/10.1145/3311927.3323150

[46] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16).* Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[47] Naja A. Mack, Dekita G. Moon Rembert, Robert Cummings, and Juan E. Gilbert. 2019. Co-Designing an Intelligent Conversational History Tutor with Children. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19).* ACM, New York, NY, USA, 482–487. https://doi.org/10.1145/3311927.3325336

[48] Joseph E. Michaelis and Bilge Mutlu. 2017. *Someone to Read with: Design of and Experiences with an In-Home Learning Companion Robot for Reading.* ACM, New York, NY, USA, 301–312. https://doi.org/10.1145/3025453.3025499

[49] Suzanne E Mol, Adriana G Bus, Maria T De Jong, and Daisy JH Smeets. 2008. Added value of dialogic parent–child book readings: A meta-analysis. *Early education and development* 19, 1 (2008), 7–26.

[50] Jack Mostow and Wei Chen. 2009. Generating Instruction Automatically for the Reading Strategy of Self-Questioning. In *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling.* IOS Press, NLD, 465–472.

[51] Michael J. Muller and Sarah Kuhn. 1993. Participatory Design. *Commun. ACM* 36, 6 (June 1993), 24–28. https://doi.org/10.1145/153571.255960

[52] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. *Patterns for How Users Overcome Obstacles in Voice User Interfaces.* Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3173574.3173580

[53] M. Naderifar, H. Goli, and Fereshteh Ghaljaie. 2017. Snowball Sampling: A Purposeful Method of Sampling in Qualitative Research.

[54] Douglas P Newton. 1995. The role of pictures in learning to read. *Educational Studies* 21, 1 (1995), 119–130.

[55] Charles A Norman and Grace Malicky. 1987. Stages in the reading development of adults. *Journal of Reading* 30, 4 (1987), 302–307.

[56] Sharon Oviatt. 1999. Ten Myths of Multimodal Interaction. *Commun. ACM* 42, 11 (Nov. 1999), 74–81. https://doi.org/10.1145/319382.319398

[57] Luiza Superti Pantoja, Kyle Diederich, Liam Crawford, and Juan Pablo Hourcade. 2019. Voice Agents Supporting High-Quality Social Play. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19).* ACM, New York, NY, USA, 314–325. https://doi.org/10.1145/3311927.3323151

[58] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (Philadelphia, Pennsylvania) *(ACL '02).* Association for Computational Linguistics, USA, 311–318. https://doi.org/10.3115/1073083.1073135

[59] Alison H Paris and Scott G Paris. 2003. Assessing narrative comprehension in young children. *Reading Research Quarterly* 38, 1 (2003), 36–76.

[60] Jackie Peck. 1989. Using storytelling to promote language and literacy development. *The Reading Teacher* 43, 2 (1989), 138–141.

[61] Juan Rebanal, Jordan Combitsis, Yuqi Tang, and Xiang 'Anthony' Chen. 2021. XAlgo: A Design Probe of Explaining Algorithms' Internal States via Question-Answering. In *26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21).* Association for Computing Machinery, New York, NY, USA, 329–339. https://doi.org/10.1145/3397481.3450676

[62] Elisa Rubegni, Rebecca Dore, Monica Landoni, and Ling Kan. 2021. "The girl who wants to fly": Exploring the role of digital technology in enhancing dialogic reading. *International Journal of Child-Computer Interaction* 30 (2021), 100239.

[63] Elisa Rubegni and Monica Landoni. 2014. Fiabot! Design and Evaluation of a Mobile Storytelling Application for Schools. In *Proceedings of the 2014 Conference on Interaction Design and Children* (Aarhus, Denmark) *(IDC '14)*. Association for Computing Machinery, New York, NY, USA, 165–174. https://doi.org/10.1145/2593968.2593979

[64] Elisa Rubegni, Monica Landoni, Laura Malinverni, and Letizia Jaccheri. 2022. Raising Awareness of Stereotyping Through Collaborative Digital Storytelling: Design for Change with and for Children. *International Journal of Human-Computer Studies* 157 (2022), 102727. https://doi.org/10.1016/j.ijhcs.2021.102727

[65] Thomas Scialom, Benjamin Piwowarski, and Jacopo Staiano. 2019. Self-Attention Architectures for Answer-Agnostic Neural Question Generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. ACL, Florence, Italy, 6027–6032. https://doi.org/10.18653/v1/P19-1604

[66] Siamak Shakeri, Cicero Nogueira dos Santos, Henghui Zhu, Patrick Ng, Feng Nan, Zhiguo Wang, Ramesh Nallapati, and Bing Xiang. 2020. End-to-End Synthetic Data Generation for Domain Adaptation of Question Answering Systems. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ACL, Online, 5445–5460. https://doi.org/10.18653/v1/2020.emnlp-main.439

[67] Timothy Shanahan and Christopher J Lonigan. 2010. The National Early Literacy Panel: A summary of the process and the report. *Educational Researcher* 39, 4 (2010), 279–285.

[68] Petr Slovák, Nikki Theofanopoulou, Alessia Cecchet, Peter Cottrell, Ferran Altarriba Bertran, Ella Dagan, Julian Childs, and Katherine Isbister. 2018. "I Just Let Him Cry...: Designing Socio-Technical Interventions in Families to Prevent Mental Health Disorders. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 160 (nov 2018), 34 pages. https://doi.org/10.1145/3274429

[69] Stacey A Storch and Grover J Whitehurst. 2002. Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental psychology* 38, 6 (2002), 934.

[70] Mingyu Sun and Joyce Y. Chai. 2006. Towards Intelligent QA Interfaces: Discourse Processing for Context Questions. In *Proceedings of the 11th International Conference on Intelligent User Interfaces* (Sydney, Australia) *(IUI '06)*. ACM, New York, NY, USA, 163–170. https://doi.org/10.1145/1111449.1111487

[71] Cristina Sylla. 2013. Designing a Tangible Interface for Collaborative Storytelling to Access 'embodiment' and Meaning Making. In *Proceedings of the 12th International Conference on Interaction Design and Children* (New York, New York, USA) *(IDC '13)*. Association for Computing Machinery, New York, NY, USA, 651–654. https://doi.org/10.1145/2485760.2485881

[72] Yumiko Tamura, Mitsuhiko Kimoto, Masahiro Shiomi, Takamasa Iio, Katsunori Shimohara, and Norihiro Hagita. 2017. Effects of a Listener Robot with Children in Storytelling. In *Proceedings of the 5th International Conference on Human Agent Interaction* (Bielefeld, Germany) *(HAI '17)*. ACM, New York, NY, USA, 35–43. https://doi.org/10.1145/3125739.3125750

[73] Duyu Tang, Nan Duan, Tao Qin, Zhao Yan, and Ming Zhou. 2017. Question answering and question generation as dual tasks. *arXiv preprint arXiv:1706.02027* (2017).

[74] Asimina Vasalou, Sara Kalantari, Natalia Kucirkova, and Yvonne Vezzoli. 2020. Designing for oral storytelling practices at home: A parental perspective. *International Journal of Child-Computer Interaction* 26 (2020), 100214.

[75] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3334480.3381069

[76] Dakuo Wang, Pattie Maes, Xiangshi Ren, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2021. *Designing AI to Work WITH or FOR People?* ACM, New York, NY, USA. https://doi.org/10.1145/3411763.3450394

[77] Tong Wang, Xingdi Yuan, and Adam Trischler. 2017. A joint model for question answering and question generation. *arXiv preprint arXiv:1706.01450* (2017).

[78] Brenda K Wiederhold. 2018. "Alexa, are you my Mom?" The role of artificial intelligence in child development.

[79] Andrew Wright. 1995. *Storytelling with children*. Oxford University.

[80] Ying Xu. 2020. Using conversational agents to foster young children's science learning from screen media. In *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*. 14–19.

[81] Ying Xu, Stacy Branham, Xinwei Deng, Penelope Collins, and Mark Warschauer. 2021. *Are Current Voice Interfaces Designed to Support Children's Language Development?* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445271

[82] Ying Xu, Dakuo Wang, Penelope Collins, Hyelim Lee, and Mark Warschauer. 2021. Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner. *Computers & Education* 161 (2021), 104059. https://doi.org/10.1016/j.compedu.2020.104059

[83] Ying Xu and Mark Warschauer. 2019. Young Children's Reading and Learning with Conversational Agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3299035

[84] Ying Xu and Mark Warschauer. 2020. Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the Interaction Design and Children Conference*. 216–228.

[85] Ying Xu and Mark Warschauer. 2020. Exploring Young Children's Engagement in Joint Reading with a Conversational Agent. In *Proceedings of the Interaction Design and Children Conference* (London, United Kingdom) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 216–228. https://doi.org/10.1145/3392063.3394417

[86] Ying Xu and Mark Warschauer. 2020. What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[87] Bingsheng Yao, Dakuo Wang, Tongshuang Wu, Tran Hoang, Branda Sun, Toby Jia-Jun Li, Mo Yu, and Ying Xu. 2021. It is AI's Turn to Ask Human a Question: Question and Answer Pair Generation for Children Storybooks in FairytaleQA Dataset. arXiv:2109.03423 [cs.CL]

[88] Xuchen Yao, Gosse Bouma, and Yi Zhang. 2012. Semantics-based question generation and implementation. *Dialogue & Discourse* 3, 2 (2012), 11–42.

[89] Andrea A Zevenbergen and Grover J Whitehurst. 2003. Dialogic reading: A shared picture book reading intervention for preschoolers. *On reading books to children: Parents and teachers* (2003), 177–200.