Teachers, Parents, and Students' Perspectives on Integrating Generative AI into Elementary Literacy Education

Ariel Han
hanjy3@uci.edu
University of California, Irvine
United States

Xiaofei Zhou
xzhou50@ur.rochester.edu
University of Rochester
United States

Zhenyao Cai
zhenyaoc@uci.edu
University of California, Irvine
United States

Shenshen Han
shenshh@uci.edu
University of California, Irvine
United States

Richard Ko
kor2@uci.edu
University of California, Irvine
United States

Seth Corrigan
scorrig1@uci.edu
University of California, Irvine
United States

Kylie Peppler
kpeppler@uci.edu
University of California, Irvine
United States

ABSTRACT

The viral launch of new generative AI (GAI) systems, such as ChatGPT and Text-to-Image (TTL) generators, sparked questions about how they can be effectively incorporated into writing education. However, it is still unclear how teachers, parents, and students perceive and suspect GAI systems in elementary school settings. We conducted a workshop with twelve families (parent-child dyads) with children ages 8-12 and interviewed sixteen teachers in order to understand each stakeholder's perspectives and opinions on GAI systems for learning and teaching writing. We found that the GAI systems could be beneficial in generating adaptable teaching materials for teachers, enhancing ideation, and providing students with personalized, timely feedback. However, there are concerns over authorship, students' agency in learning, and uncertainty concerning bias and misinformation. In this article, we discuss design strategies to mitigate these constraints by implementing an adults-oversight system, balancing AI-role allocation, and facilitating customization to enhance students' agency over writing projects.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

Generative AI, Artificial Intelligence, K-12 Education

ACM Reference Format:
Ariel Han, Xiaofei Zhou, Zhenyao Cai, Shenshen Han, Richard Ko, Seth Corrigan, and Kylie Peppler. 2024. Teachers, Parents, and Students' Perspectives on Integrating Generative AI into Elementary Literacy Education. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3613904.3642438

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642438

1 INTRODUCTION

In early January 2023, The New York Education Department announced a ban on using generative AI chatbots (ChatGPT) in school districts' networks and devices over concerns about potential misuse and safety [90]. By May of that year, however, the department dropped the ban, announcing plans to explore whether there were potential possibilities to use the technology in the classroom [60, 91]. When new technology is introduced in educational settings, perceptions often swing between excessive optimism and skepticism, largely due to the uncertainty surrounding the actual usage of these systems in real-world scenarios [27, 87]. The ongoing discourse in education around generative AI (GAI) emphasizes the need for comprehensive research into its integration within educational contexts [98].

GAI, also known as Generative Adversarial Networks (GANs), GAI systems have gained significant attention within the HCI community [75, 109, 124]. The advances of generative AI (i.e., ChatGPT, Dall.e 2, Midjourney) open up a new horizon of open-context conversation with an AI chatbot [4, 20, 74, 78], including generating novel outputs-such as images, text, music, or video-based on patterns it learned from large datasets during its training [15]. The HCI research community has started to examine utilities and interaction techniques with these systems [63, 116], focusing on new interaction styles [52, 125], Large Language Models' (LLM) capacities [63], and how to adapt the systems to creative activities for adults [43, 68]. While the advancements in GAI have captivated the HCI community with their ability to foster novel forms of open-context interaction, applying these technologies in educational settings, especially for elementary school students, presents a different set of challenges and opportunities.

Technology integration in education requires understanding practical realities rather than relying solely on technological advancements, which call for balanced approaches that recognize the complexities of teaching and learning [87]. Recognizing the role

of storytelling in child development [53] and its impact on critical skills like imagination and comprehension [69], it becomes clear that integrating such advanced technologies in education demands a careful balance. This approach should respect both the potential of GAI and the intricate nature of teaching and learning processes, ensuring that technological advancements are meaningfully and effectively aligned with educational needs and realities. Considering the need to underscore the applicability of leveraging GAI in writing instruction for students, we conducted a study to examine the different perspectives of stakeholders in K-6 education (i.e., teachers, parents, and students) regarding the integration of GAI in elementary school literacy education. Our objective is to understand stakeholders' aspirations and concerns regarding the use of new systems in academic settings in a holistic manner by including both teachers and learners so that the HCI research community can use these insights to design and develop GAI-powered educational applications that are safe and productive for elementary school students writing.

In this study, we sought to answer the following questions:

- How do stakeholders in elementary school settings–parents, teachers, and students–perceive AI to support teaching and learning writing projects, and what are their opinions of the potential benefits and limitations of leveraging it? How do values and motivations towards GAI systems differ among stakeholders in education?
- In what ways can GAI systems be designed so that they are effective, engaging, and safe for teaching literacy for 2nd to 6th graders?

To answer these questions, we conducted workshops with families with children ages 8-12 (i.e., in 2nd through 6th grade) that included semi-structured interviews with students and parents during and after the workshop. Also, we carried out 1:1 semi-structured interviews with 16 teachers to better understand teachers' motivations, perspectives, and strategies for leveraging GAI in writing projects. In total, we report on insights from 40 participants who present unique perspectives on GAI from three groups of stakeholders in education (i.e., 16 teachers, 12 parents, and 12 students).

From the study, stakeholders' perceptions towards GAI systems and their opinions of potential benefits and challenges related to writing surfaced three major themes: 1) teachers' view as a part of *digital citizenship development*, 2) parents' perception of new types of *toys, games, and screen time*, and 3) students' perceptions as *smart and helpful companions*. In addition to these major themes, we highlight possible obstacles and concerns regarding authorship and ownership issues over writing outputs, challenges examining students' agency in learning, and difficulties in controlling bias and hallucinated content created by GAI systems. Based on the findings, we provide design implications to mitigate the shortcomings of these systems in educational settings. This discussion includes: 1) navigating the complexity of authorship in AI-assisted writing systems through examining a child-AI interaction chatlog, 2) enhancing student agency through role allocation and curating AI personas in GAI systems to promote independent writing and cultivating conversations aimed at fostering students' unique voices, and 3) balancing flexibility and control with teacher-in-the-loop GAI-LLM systems that allow teachers to curate child-AI interaction. We aim to contribute to the HCI community by highlighting the practical applications and limitations of GAI in education and by offering insights that can guide the design and implementation of GAI tools in a way that aligns with the needs and concerns of various educational stakeholders.

Two main contributions are made by this work:

- Our study provides a qualitative investigation of the efficacy of generative AI for writing projects, surfacing potential benefits and challenges in using LLM-driven chatbots in educational settings. Our findings demonstrate that GAI systems offer opportunities for creating adaptive teaching materials tailored to students' unique competencies in writing, broaden ideation and timely interaction through dynamically generated learning resources, and provide individual, culturally relevant feedback. At the same time, using GAI systems in writing carries significant limitations regarding authorship, agency, and potential misinformation.
- We present design implications by investigating ways to harness generative AI in writing projects safely and effectively. We surface the challenges and difficulties from stakeholders' perspectives and provide insight into designing new systems. We propose design suggestions to enhance safety by balancing flexibility and control through teacher-in-the-loop systems where teachers can prompt to curate AI agent capacity with prompt bank interfaces, designing the AI agent persona as coach or/peer rather than an assistant, and designing role-allocation among AI and students of which students have the freedom to write independently, edit, customize themselves instead of having the AI agent generate on their behalf.

2 RELATED WORKS

In this section, we examine research literature related to the implications of artificial intelligence for education in HCI research, as well as educational research related to artificial intelligence applications for learning and teaching in educational settings.

2.1 Tracing the Evolution of Technology in Education: Implications for Modern AI Integration

Reflecting on the past usage and integration of new educational technologies in real-world educational settings can offer valuable insights for predicting and enhancing their effectiveness in learning environments [87]. To contextualize our investigation of the potential applications and benefits of emerging GAI systems in educational contexts, we trace the impact of Massive Open Online Courses (MOOCs) and Intelligent Tutoring Systems (ITS). These technologies have been pivotal developments in the history of scalable learning with implications for the educational sector. Despite rapid technological advancements, the anticipated radical transformation in education by innovative educational technology companies (e.g., Khan Academy, Udacity) has largely fallen short of expectations. Personalized learning platforms claim to tailor education to individual student needs, but they often fall short in practice due to the complexities of learning processes, effective pedagogies,

and the constraints of algorithmic customization [87]. Therefore, Reich 2020 argues that educational innovations must be deeply rooted in the realities of teaching and learning.

Reich [87]'s four dilemmas highlight the complexities of learning at scale platforms, emphasizing the need for a critical reassessment in the context of emerging Generative AI (GAI) technologies. These dilemmas include the preference for familiar tools, the unequal benefits of new technologies, the challenge of nuanced assessment beyond binary right or wrong answers, and the issues of data privacy and equity [66]. As GAI systems offer more natural and adaptable human-AI interactions, they present an opportunity to address these challenges, making AI-based educational tools more accessible and equitable for diverse learners.

Recent advancements in Large Language Models (LLMs) enhance their ability to assess human reasoning in writing, moving beyond the traditional right-or-wrong evaluation methods of current Intelligent Tutoring Systems (ITS). This progress offers a more nuanced understanding of student logic and thinking, enabling personalized and adaptive feedback. Studies, such as those by Steiss et al. [98], are beginning to explore GAI's potential in analyzing and understanding the nuances of students' written work and reasoning processes, which pose potential capabilities to integrate algorithmic guided instructions flexibly.

2.1.1 Integrating Artificial Intelligence in Education. The use of artificial intelligence in education (AIED) has been explored through the application of intelligent tutoring systems, conversational agents (CA), and chatbots. These technologies have enhanced teaching and learning [11, 21, 55, 76, 84, 101, 117, 121], yet little of this prior work addresses directly how AIED integrates holistically into educational settings [21, 54]. An exception to this is Chiu et al. [22] systematic review of AI's roles in education, which surfaces potential benefits of AI for learning, including providing adaptive learning by assigning tasks based on individual abilities that enhance academic performance and facilitating human-machine conversation to motivate and engage students. However, Chiu. [22] pointed out the need for further studies that examine students' educational outcomes with AI-based systems (such as chatbots or conversational AI).

The HCI community has also provided insights into the perception of AI systems [106] among educational stakeholders, including teachers [65, 85], children [6, 16, 115, 118, 122], and parents [39, 40, 105, 120]. To design AI tools and curriculums that align with the values and contexts of stakeholders in education (i.e., teachers, parents, and students), Lin and Brummelen. [65] conducted co-design workshops with K-12 teachers to develop design recommendations for creating AI curriculums and tools aligned with teachers' needs. Their findings revealed how teachers value learning outcomes, student engagement, ease of use, and collaboration when incorporating AI in the classroom. Design recommendations from the study emphasize the importance of designing AI tools to be adaptable to diverse contexts (e.g., different grades and subjects).

Outside the classroom, parents see technology (including AI) as a way to enhance parent-child interactions by selecting content for their children, showing a preference for customized content [127]. Children's views on AI agents differ based on age and their performance in AI experience and interaction. Younger children often perceive AI agents as intelligent toys, while their older counterparts perceive them more as humanoid entities with lesser intelligence [115]. Additionally, Xu and Warschauer. [122] reported that most children view conversational agents (CAs) as having cognitive capabilities via continuous communication but possess fewer psychological entities (i.e., having emotion). The findings suggest possibilities of designing CA as a learning companion, incorporating social interaction and emotional feedback [122].

Despite this body of recent research, there is still a lack of clarity regarding the role of artificial intelligence (including generative AI) from educators, parents, and students' standpoints. Additionally, further research is needed to investigate whether and how these emerging technologies can improve the learning process of literacy development in elementary school settings.

2.1.2 Emerging Trends and Challenges in Generative AI Applications for Education. The rapid advancement of GAI, such as large language models (LLMs) and Text-to-Image (TTI), learn patterns and structures from existing data and generate new content [113]. These breakthroughs have led to a new generation of dialog systems that enable the possibility of leveraging the system to facilitate open-ended discussion and generate educational content for teaching and learning [79]. Ahmad et al. [2] examined the implications of ChatGPT in the education sector, emphasizing the need to develop skills for using LLMs and GAI to be prepared for future job markets. This requires students to know how to prompt AI systems effectively and to be able to analyze the quality, originality, and accuracy of the results [2].

Research on AI systems in literacy education (reading and writing) focuses on LLM-based chatbots for language learning [1, 128], scientific writing [42], creative writing [24, 97, 126], and creative storytelling [50, 127]. For example, Gero et al. [42] studied how LLM-powered co-writing platforms can enhance engagement and idea generation with STEM graduate students. Yuan et al. [126] studied adult hobbyist writers' sense of ownership over AI-assisted writing and found that AI integration does not undermine writers' feeling of ownership because writers use AI-generated text as an inspiration rather than taking it verbatim. Lee at al. [63] also conducted studies with adult participants to understand the affordance of large language models (LMs). The authors aimed to guide the design of LLM applications and developed a CoAuthor system, which focuses on capturing and analyzing user engagement data. This system tracks how users collaborate and construct stories, providing valuable information on user interactions and narrative development within the context of LLM applications. The findings showed that CoAuthor enhances writing productivity, increasing the text writers produce. But Yuan et al. [126] and Lee et al. [63] also raised questions about writer's feeling of ownership over their writing outputs and indicated the results were uncertain.

Recent GAI-powered educational applications offer potential opportunities to leverage GAI systems in teaching and learning (GPT-3, TTL) [17]. For example, Speak [5] uses GAI systems (GPT-3) to simulate smooth verbal conversation with learners to improve English speaking proficiency without age limit. Also, web applications and conversational agents (CA) have been developed to support students' reading comprehension through story creation (i.e., Wanderly, OnceUponABot, AlexaBedtimeStory) mainly for families with children ages 5-12 [14, 71, 96]. MagicSchool.ai [3]

is a web application that uses GAI systems to support efficient lesson plans for teachers by suggesting and generating quizzes and scaffolded lesson materials. Khan Academy recently launched an LLM-based AI agent, Khanmigo, that carries a text-based conversation with students as a tutor, as well as facilitating teachers' versions as teaching assistants, which assist teachers in creating lesson plans for a wide range of subjects (history, language arts, math, foreign language) across K-12 [61].

However, it is still unclear how these new interactions, user experiences, and learning engagements affect learning outcomes [10]. The current story creation apps powered by GAI systems produce whole stories for students, which raises a question about whether it could promote language learning or undermine creativity [50]. Hence, further research is needed to ensure GAI-powered learning tools are effective and age-appropriate.

2.2 Child-AI interactive systems

Nowadays, an increasing number of children interact with AI-enhanced products daily. Researchers have explored the perspectives of various stakeholders, including teachers, parents, and children. Findings reveal that parents desire CA to foster children's social engagement and involve parents in in-home learning [41]. However, researchers raised concerns about the lack of open-ended and extended back-and-forth dialogue while considering CA to support children's language development [119]. Enhancements are also needed for human-AI collaboration to relieve the repair burden on families during their communication breakdowns with CA [11]. As for children's perspective, research efforts have been made to investigate children's perception of their data utilized online [110], children's autonomy over the technology [111], and AI technologies' influence on child development. With the recent advancement of GAI, such as LLM, daily life AI-enhanced products have largely extended their power of human-AI collaboration, including children-AI co-creation. This is also aligned with the rising desire for AI literacy education outside of the computing domains [94], and in turn, challenges AI literacy education by equipping children with some basic AI literacy in both classroom and family scenarios [33, 99]. These all require a deeper understanding of stakeholders' needs and concerns around child-AI co-creation.

Existing child-AI co-creative systems encompass interactive storytelling [126], creative writing [35], and drawing [127]. Wordcraft [126] is a text editor fostering collaborative engagement between users and LLM in storytelling. It facilitates open-ended conversations related to the narrative, responds to users' natural language queries, and offers suggestions to assist writers in overcoming creative hurdles. The study with adult participants suggests incorporating real-time requests and predefined controls to amplify the co-creative experience. In the intersection of drawing and creative storytelling, "StoryDrawer" aims to support children in creating oral stories during visually immersive storytelling episodes [127]. Results from the evaluation with children highlight the importance of encouraging collaboration and co-creation between children and the AI system rather than solely relying on the system to generate stories. CreativeBot is a robot designed to stimulate children's creativity through co-creative storytelling [35]. The robot's ability to generate unexpected and surprising story elements proved particularly effective. Findings imply flexibility, adaptability, collaboration, and surprise as crucial factors for the CreativeBot. Besides such conversational, drawing, or robotic interactions, researchers have developed different LLMs as supports for collaborative creative writing [77, 100], where creativity requires writing with a relevant purpose, understanding, judgment, and evaluative abilities in ways that are deemed original and valuable to a community [26]. However, by this definition, by relying primarily on summation, LLMs lack the intention to write and do not possess the self-feedback loop necessary to intentionally deviate from conventions, hindering their capacity [38]. Therefore, specific interface elements need to be designed to compensate for such limitations of LLMs. Beyond such inspection from the technical perspective, research is needed to develop a more in-depth understanding of children's, parents', and teachers' needs and concerns around child-AI co-creative systems.

3 METHODOLOGY

In the previous sections, we outlined the adoption of AI technologies in education and addressed a research gap resulting from a lack of holistic understanding of teachers, learners, and their caregivers' opinions. Considering potential stakeholders' perceptions of GAI systems may provide design implications to help guide the development of GAI tools and systems for elementary school students. To elicit stakeholders' perspectives on potential possibilities and limitations of a GAI-LLM chatbot system for writing, we conducted a workshop with families with children ages 8 to 12 (parents, N=12 and children N=12) that focused on how they used a text-to-image generator (i.e., Stable Diffusion [31]) and a chatbot powered by LLM (i.e., ChatGPT [19]). Following the workshop, we conducted 1:1 interviews with teachers who specialized in teaching writing in elementary school settings (N=16). In total, we reported on insights from 40 participants who interacted with both tools. Participants were recruited from our researcher's network (mailing list and contacts) and snowball sampling. We sought to identify teachers, parents, and students' motivations, challenges, and opinions with the new systems, elicit their concerns, and identify their perceptions and strategies in writing using GAI platforms.

3.1 Study Procedure

3.1.1 Workshop with Families. In April 2023, we conducted a workshop with families with children ages 8 to 12 (2nd and 6th graders) (Table 2) in order to better understand students' strategies and struggles when interacting with the current state of LLM-based chatbots and text-to-image generators. Their parents' and guardians' opinions and perceptions regarding using the systems for writing projects were also considered. We focus on the 8 to 12 age group, recognizing the critical importance of this phase in developing reading and writing skills [48]. This period is pivotal as children transition from learning to reading to reading to learning, a fundamental shift highlighted in Loveless's 2023 article. Early interventions during this stage can greatly influence a child's educational path and future opportunities[47]. Given this, our study aims to investigate how enhanced engagement with writing and literacy activities facilitated by GAI platforms can positively or negatively impact learning in these formative years.

Table 1: Workshop schedule

Time	Activity
15 min	Introduction (Icebreaking)
15 min	What is AI? (Discussion)
15 min	Let's learn Generative AI
15 min	Break
15 min	Let's learn GAI tools (ChatGPT and Stable Diffusion)
30 min	Let's use generative AI to write a visual story
15 min	Share your story (Reflection)

Table 2: Participants' information for the family workshop

Alias for parents	Age	Alias for students	Age
P1	37	S1	9
P2	39	S2	10
P3	36	S3	10
P4	37	S4	9
P5	42	S5	8
P6	40	S6	9
P7	36	S7	8
P9	49	S8	10
P10	41	S9	12
P11	45	S10	12
P12	38	S11	11
		S12	10

Parent participants (n=12) completed a screening survey before the workshop to ensure they were 18 or older and lived with children ages between 8 to 12 years old. The average age of parent participants was 39.8 years old at the time of the workshop, of whom (10/12) were female and (2/12) were male. According to parent reports, the mean age of the student participants was 9.8 years old, and (5/12) were girls. Eleven children (11/12) were identified as Asian American, four children (4/12) spoke only English at home, and the remainder were bilingual (6/12) or spoke English as a second language (2/12). All children possessed sufficient oral English proficiency for daily conversation. The median household income of the twelve families is $118,749, with a range from a minimum of $29,999 to a maximum of $200,000. Given the socioeconomic standards of the West Coast, USA, this income bracket is typically classified as upper-middle class [108]. It was the first time the students had used GAI-LLM chatbot and Text-to-Image generators (TTL), while seven parents (7/12) reported already using them. Family participants were compensated $25 for their time and effort.

The 2-hour, 1-day workshop was conducted in a library in a Southern California metropolitan city. Accompanied by their parents, children were required to create a visual story using a text-to-image generator (i.e., Stable Diffusion) and a chatbot powered by LLM (i.e., ChatGPT). During the writing project, we sought to understand the students' strategies and their interactions with the system through observation by taking field notes and voice recording youths' verbal expressions and semi-structured interviews [88]. Given the California State Standards in elementary literacy education, we chose narrative writing activities for students [46] instead of giving students a specific topic to write about; students wrote creatively without limitations. The topic of the visual story was open-ended, and students picked a topic based on their own interests. To assist, several prompt examples were provided (e.g., "I would like to write a topic of the story, how can I start?", "Can you list five story ideas?") before they began writing. Students worked individually without their parents' intervention unless they needed to access a required platform (i.e., Google Classroom, Google Docs). Students used the systems under the supervision of researchers.

We created a Google Classroom for the workshop that served as an information resource as well as a repository for participants' finished visual stories. Students were allowed to use the Text-to-Image generator and LLM chatbot to develop their stories. One of the researchers ran the workshop, and the other researcher observed, took field notes and conducted semi-structured interviews with children during and after the workshop. While students worked on generating their visual story, one of the researchers conducted semi-structured interviews with the parents. Interviews were recorded using a voice recorder, no videos were taken during the workshop, but pictures were taken, and students' artifacts were collected. We sought to understand students' opinions and their perceptions of AI by posing the following questions [88]: What do you like or dislike about using ChatGPT and Stable Diffusion for your creative writing and visuals?, Have you found AI useful?, How can artificial intelligence help you? To understand parent's opinions and their perspectives on using the systems for their children, one of the researchers conducted semi-structured interviews with the parents while students worked on creating visual stories. With parents, we discussed the following topics: How do you think AI impacts your child's learning?, Do you want your child to use AI or learn about AI?, What is your overall impression of using AI for your child? Interviews were recorded using a voice recorder, no videos were taken during the workshop, but pictures were taken and students' artifacts were collected.

3.1.2 Teachers' interviews. Teacher interview data collection was conducted online between June to August 2023. Teachers were recruited using similar snowball recruitment efforts as the families, with the only criteria for eligibility being that they were either current or former K-12 teachers. The teachers we interviewed (n = 16) were elementary classroom teachers from 1st to 7th grades, most of whom (14/16) work in public schools. Thirteen teachers (13/16) specialized in teaching writing and were affiliated with the National Writing Project (NWP) network. Teaching experience averaged 13.3 years (min=1.7 years, max=32 years). More detailed participant information can be found in Table 3. The majority of the teachers (8/16) are located in the United States (California and Pennsylvania), and four of them are in Asia (South Korea and China). The majority of the teachers (14/16) work in public schools, with only two working at private schools (see Table 3).

The teacher interviews were conducted individually for up to an hour via video conferencing due to geographical distances, with an average length of approximately one hour. We sought to elicit their current teaching practices, struggles, and motivations when teaching writing to their students. Afterward, we introduced GAI systems (i.e., features and functionalities) and asked about their

Table 3: Participants' information for the interview study

Alias	Grade Taught	Years of teaching	Location
T1	2nd grade	3 years	Nanjing, China
T2	6th grade	2 years	Pennsylvania, USA
T3	3rd grade	5.8 years	Incheon, S.Korea
T4	1st grade	32 years	California, USA
T5	5th grade	30 years	California, USA
T6	3rd grade	13 years	California, USA
T7	2nd grade	24 years	California, USA
T8	6th grade	14 years	California, USA
T9	5th grade	15 years	California, USA
T10	8th grade	20 years	California, USA
T11	6th grade	5.2 years	Seoul, S.Korea
T12	7th grade	1.7 years	California, USA
T13	6th grade	5.3 years	Seoul, S.Korea
T14	5th grade	6 years	Seoul, S.Korea
T15	6th grade	3.4 years	Seoul, S.Korea
T16	5th grade	12.8 years	California, USA

experiences and opinions about adapting them in educational settings specific to writing activities with their students. Most teachers (10/16) already have experience with ChatGPT and relevant GAI systems (Midjourney), whereas the rest were unfamiliar with these systems. Teachers who have used GAI systems continue to use it in their teaching practices since they first tried, and their years of teaching experience are averaged at 5.8 years, compared with 22.5 years for teachers who have never used GAI systems.

In the interviews, the following topics were discussed:

- The interviewee's general practices, difficulties in teaching, and concerns for their students (e.g., "What is the hardest part in teaching writing in your class?"),
- Their experiences the state-of-the-art GAI systems (i.e., ChatGPT, Stable Diffusion) (e.g., "What is your level of familiarity with Generative AI systems like ChatGPT and Stable Diffusion? "Have you ever used or willing to use the GAI systems in your class or for yourself?"),
- Their opinions of their intended usage of the GAI systems, and their opinions and concerns about them (e.g., "Can you tell me your thoughts about the GAI systems as students use them for writing?", "Can you share your opinions on whether or not GAI systems are beneficial or harmful for students?", "How do you envision these systems being used by teachers or students?")

A recording of all interviews was conducted with the consent of the participants, and teacher participants were compensated $25 for their time and effort. Our study was approved by the authors' institutions' institutional review boards (IRBs).

3.2 Data Analysis

The interview data was first transcribed using an automatic transcription program (Otter.ai) that maintained the original audio and aligned it with the transcript. After thoroughly reviewing the transcript, we transferred the transcript to a qualitative data analysis software (Atlas.ti) ensuring that the original audio was preserved and accurately aligned with the transcripts. Following this, we utilized qualitative data analysis software for an initial round of open coding, adhering to established qualitative research methodologies [93, 103]. We conducted an inductive approach to analyze interview data [102]. Following the inductive approach, two researchers independently read the transcripts and identified key themes and patterns within the text. This collaborative and iterative process of theme identification and analysis was instrumental in reaching theoretical saturation [70]. Each researcher assigned the first round of low-level codes guided by our research questions (e.g., participants' opinions (stance) of the potential benefits and limitations of leveraging GAI; how their values and motivations differ) into each theme. In order to reduce overlap between themes, we repeated discussions with researchers. We categorized the low-level codes into higher-level themes. The researchers regularly discussed (every week for two months for an hour each) and iterated to construct the themes. By systematically coding the data and constantly comparing emerging themes, we were able to ascertain when no new themes were emerging from the data, indicating that theoretical saturation had been achieved. We organized our results around the main theme of the advantages and challenges of using LLM chatbots for educational purposes in K-6 settings, which emerged from this coding. We categorized codes into four high-level themes (i.e., perception, positive opinions, negative opinions, and suggestions). The analysis contained nine mid-level themes (i.e., teachers' perception of digital literacy development, parents' perception of toys and games, students' perception as helpful companions, creating adaptive teaching content, timely interaction and broadening ideation, personalized and culturally relevant feedback, lack of context for students, problems with authenticity and authorship, hard to distinguish students' agency, difficult to control biased and misinformation) and 34 codes under each theme.

3.3 Limitations

Our study focused on the context of educators and families in one of the metropolitan cities on the West Coast, United States, as well as mid-high socioeconomic families. It is possible that our findings do not represent the perspectives of all populations on LLM-based education chatbots for writing. Additionally, the majority of families in the study were multilingual, primarily Asian-immigrated families (7/12) whose children were born on the West Coast of the United States and attended public schools. Since our samples lack a diverse cultural background, some of their perspectives and opinions might be limited. The majority of parent participants were mothers (11/12), and eight mothers (8/12) were stay-at-home with an average age (of 39 years old); hence, their views and opinions from the interviews are hard to represent all parents' perspectives towards GAI systems for their children's writing project. Additionally, considering the majority of teachers we interviewed are from high-SES school districts, their teaching practices, motivations, and concerns are likely to differ from those of other teachers, so generalizing their views is problematic. A future study should also consider interviewing school district administrators, whose voices are central to systemwide policy decisions.

Additionally, during the workshop, we missed the opportunity to collect chat logs to investigate students' interaction techniques

with a chatbot. Similarly, while we reviewed the final output of the students' writing pieces, it would have been better to check the history of their editions in Google Docs in order to understand their contribution to the writing better, whether they simply copied and pasted from AI-generated text, or how much they wrote by themselves. An analysis of the student's perception of ownership and the actual percentage of contribution to the piece would be valuable, as well. It may be worthwhile to investigate in the future if there are different ways to assess and measure students' learning in AI-students co-writing projects in the classroom.

4 FINDINGS

By analyzing qualitative interviews and observational notes, we uncovered multiple perspectives regarding the use of GAI in literacy education. In this section, we report major findings regarding our participants' opinions and experiences with GAI. We outline the values and perceptions of multiple stakeholders (see Figure 1), then elaborate on the findings in the advantages and constraints of GAI for literacy education (see Figure 2). The findings are categorized by each stakeholder's viewpoint to highlight how their values and perspectives differ. Following that, we categorize the themes into teaching and learning and integrated stakeholders' opinions, as stakeholders often have insight into other stakeholder perspectives (e.g., teachers' perspectives on students; and parents' perspectives on their children).

We report major themes in our stakeholders' perspectives and opinions about using GAI in literacy education, particularly teaching and learning writing. GAI is perceived differently by each stakeholder, including 1) teachers' view as a part of digital citizenship development, 2) parents' perception as new types of toys, games, and screen time, and 3) students' perceptions as smart and helpful companions.

4.1 Multifaceted Views on the Role of GAI in Literacy Education

4.1.1 Adapting Digital Transformation: Teachers' Perspective on Integrating GAI in Digital Literacy Development. Our results indicate that teachers acknowledge that their students will grow up in a society where emergent digital technology is an integral part of life. Nine teacher participants (9/16) expressed willingness to promote the use of GAI to foster safer and healthier ways of using the systems. Specifically, T3 noted:

"I do think that instead of rejecting it, we need to figure out how it works for us and what we need to do with it. I mean, our students are going to be using it, our co-workers are going to be using it, right? It's going to be in the world. So I do think we're better off to figure it out than to reject it for sure."

While over half of teachers tried to embrace the GAI systems into their practices, (7/16) considered them as an essential part of the digital citizenship development for both teachers and students, agreeing to teach students about GAI systems as another tool that they will need to learn how to use.

Teachers pointed out that GAI systems can also be used to support educational processes [23], nine respondents (9/16) emphasized that GAI systems like ChatGPT and Text-to-Image generators can be integrated into their instructional processes:

"I think it has a lot of potential. I think there's lots of excitement for potential teachers in lesson planning. I don't think it's kind of replacing any existing curricula. But I think it can be a tool to extend the teaching as a part of the process."

For instance, one respondent noted that the current GAI–LLM chatbot lacks the capacity to be fully integrated into human conversations but can be useful for brainstorming ideas:

"I don't think AI has been adapted to fully understanding or answering questions yet, but I have used it a ton as a student and a professional to brainstorm ideas. It's like a friend with a wealth of information, like someone I bounce back ideas from."

Our findings indicate that teachers are willing to integrate new systems (GAI) into their teaching pipeline along with digital literacy development. In addition, they stressed the importance of equipping their students with the ability to use GAI systems to develop their digital citizenship.

4.1.2 Parental Caution: Attitudes Toward GAI Systems in Children's Literacy Education. On the whole, parents expressed more conservative attitudes, with (11/12) of respondents expressing skepticism about the use of GAI systems in their children's education. In spite of the fact that all participants in the parents' interview (12/12) agreed that AI will be a part of their children's lives as they grow, it is still important to know how to use it properly. For parents of children ages 8 to 12 years old, it is more important for their children to learn how to use GAIs responsibly and safely, which makes them more cautious about potential harm.

Seven parents expressed concern over uncertainty and data privacy when their children played games or watched videos with real-time chats with anonymous strangers on the internet; they found AI such as Alexa or Google Play to be safer. According to P01,

"My kids also play with Anonymous. I'm so worried because of the anonymous player, we don't know if the person is good or bad. So, if my kids are going to play with anonymous players, I would choose to play with AI because I think AI is at least safer than those harmful people."

Also, we identified a conflict between their values and their perception of GAI systems. It is important for parents to prioritize their children's overall well-being and well-rounded development (i.e., soft skills, emotional, physical, and intellectual), not just hard skills and academic success (i.e., test scores and grades). Eight parents (8/12) emphasized their focus on literacy education and their willingness to support it through child-centered approaches and interest-driven experiences (e.g., purchasing books their children are interested in reading). However, these parents perceive GAI systems like ChatGPT and Text-to-image generators (TTL) as other types of games and toys that will increase their children's screen time. P1 said:

"I mean, for kids, ChatGPT and Stable diffusion are just another type of toys. It's like they play Roblox or Minecraft or AI graphics."

There also appeared to be a generation gap between parents and children over AI perception, mirroring the lack of confidence for parents to introduce new technology to their children that has existed for decades [86]. Most parents (8/12) perceived the GAI systems as new to them, so they had difficulty imagining how it would affect young minds. For example, P03 and P04 mentioned:

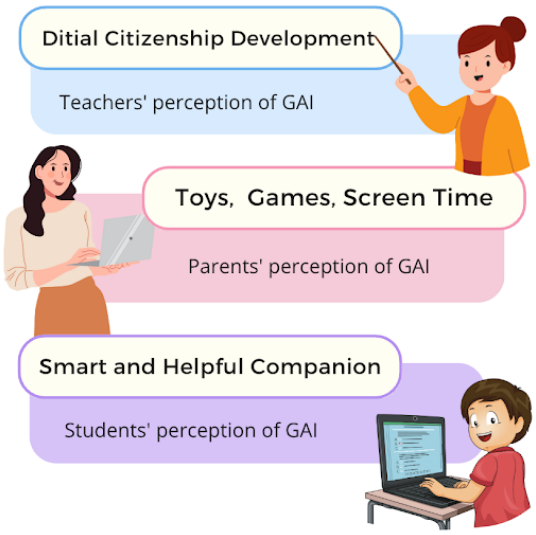

Figure 1: Summary of each stakeholder's perspectives and opinions of GAI systems (top: teachers, middle: parents, bottom: students).

"I have no idea. Because I don't know AI exactly, Because I didn't learn it when we were young, it's hard to say it's unnecessary because we don't know it well. That's the problem. So the parents like us from the generation that we don't even have AI."

While such expressions of distrust are rooted in a lack of knowledge and experience, some parents identified that learning the new system with their kids could serve as a learning opportunity for them both. P08 highlights,

"So things are maybe an opportunity for parents to learn with a kid at some time. Okay, so they get to know what AI is like and how to use AI."

As such, even though all parents acknowledged that their children need to learn how to use GAI systems properly, most parents prioritized promoting critical thinking and problem-solving instead of introducing GAI systems to their children. Moreover, parents (n=11, mothers) presented anxiety over adapting GAI systems for their children's writing projects, which could limit their children's creative thinking. Hence, they were curious about finding a way to leverage GAI systems for themselves as adults and using it for their children instead of directly giving them to their kids (i.e., creating word quizzes for their children).

4.1.3 Creative Allies with Caveats: Students' Mixed Perceptions of GAI Systems in Literacy Projects. For students, data from the workshop revealed that they (9/12) regard chatbots and TTLs as creative, smart, and helpful companions in the process of creative visual story writing, as S1 mentioned:

"I initially thought that artificial intelligence wouldn't be able to do creative things because it doesn't have a brain or mind, but it turned out more diverse and creative than I expected, which surprised me."

The vast majority of students (11/12) were optimistic about using the GAI in the process of creative writing, with (10/12) of students pointing out the efficiency of using the GAI-LLM chatbot and TTL generator to enable rapid prototypes, which broadened their choice of ideations. S7 highlighted,

"I can use this to test out as many as my ideas. I think it's really efficient."

We observed two primary difficulties encountered by students when they started the systems: 1) initial user prompts and 2) deficient AI responses. Many students had difficulty figuring out what to do due to the blank interfaces and lack of instruction and context on the website. Once we provided guidance on how to start (i.e., an example prompt included "Can you generate five story ideas for a children's book?"), they began testing them and learning how to use the system. Half (6/12) of the students also complained at times that GAI had not generated the content they intended. As a result, we concluded that instructing and teaching prompt writing would enhance efficiency and adaptability [67]. Second, we found that the randomness of the output generated by GAI systems can be a double-edged sword. Despite the possibility of unexpected, sometimes inappropriate results (e.g., generating a dead animal), Seven students (7/12) saw these moments as chances to expand their ideation, as they are likely to view even unexpected outcomes as part of the divergent process of their conception.

4.2 Delineating Advantages: GAI's Contributions to Literacy Education

To elaborate on the findings about the advantages of GAI in literacy education, we categorized the themes from our interviews and observations into teaching and learning aspects. In each section, all stakeholders' perspectives are incorporated since stakeholder perspectives represent other stakeholder perspectives (parents concerned about their kids' privacy, teachers' views about their students). Findings demonstrate that the advantages in teaching include enhancing efficiency in teaching by enabling fast and easy construction of scaffolded materials and content, including pre-instruction (by developing different levels of materials tailored to each student's abilities), during instruction (by facilitating questions and quizzes), and post-instruction (by developing a rubric). In

Figure 2: Summary of our findings of potential affordances and limitations of GAI systems for writing projects in elementary school settings

terms of how this affects user learning, GAI enables personalized experiences that provide immediate feedback to support the needs of diverse learners (i.e., by facilitating a real-time GAI-powered tutoring system). Further, interacting with GAI encourages students to generate ideas around topics, add details, and apply culturally relevant approaches (see Figure 2).

4.2.1 Enhancing Pedagogical Efficiency: GAI in Crafting Customized and Scaffolded Mentor Texts. The teachers (16/16) all affirmed that GAI systems can be used to create adaptive teaching materials as part of their lesson planning. In particular, the majority of the teachers (13/16) who specialized in writing education highlighted the potential for GAI systems to generate scaffolded mentor texts (i.e., texts that model for students what good writers do) that allow students to adapt and learn from the authors' writing style (i.e., words, sentences, or paragraphs). T7 highlighted,

"A lot of the craft of writing comes from looking at examples and finding out what the experts did and using what we've learned in our own pieces. Let's say we've studied this particular sentence deeply, and then we won't just imitate it; we find it out on our own and then try it on. Then, the kids change that for themselves. I use a ton of mentor texts."

However, nine teachers (9/16) pointed out the difficulties of finding and incorporating mentor texts that can be seamlessly integrated into their curriculum at the appropriate level for all students. T3 mentioned,

"Using mentor text is really a lot of teacher work to design it and figure it out. And what if I could generate mentor sentences and have everything ready to go. I would love that. That is one of the ways that we can use it to help us develop some of the mentor texts that we have spent hours looking for."

The elementary school classroom teachers (14/16) stated that their students have different literacy levels and interests, so a standardized curriculum makes it hard to tailor learning materials to each student's unique abilities. In response, teachers imagined leveraging GAI systems like ChatGPT to generate scaffold vocabularies and sentence levels tailored to each student's unique level. According to T13,

"Can I use Generative AI to develop reading materials at different levels for kids to read? I would love to be able to put in a topic and get information coming out, such as climate change. What would be even much better if you could layer on phonics? I can now do phonics instruction and help support within the realm of the science of reading. Having such a tool would be a tremendous time-saver, simplifying the lengthy process of sourcing and summarizing appropriate materials for diverse classroom needs."

Other teachers emphasized that they can use GAI systems to generate mentor texts because they can evaluate the quality of the texts and ensure the content is accurate. As one instructor pointed out, teachers are able to determine whether the GAI-generated content is appropriate or not. As T6 pointed out,

"Recently, I used Generative AI to create a mentor text, saving a lot of time. Since teachers have a solid understanding of the topic, we can verify the facts and integrate them into our teaching process. There's definite learning potential in this approach."

This implies the potential opportunity for teachers to use the GAI systems to generate scaffolded mentor texts and teaching materials for different levels of students' capacity.

4.2.2 Scaling Individual Attention: GAI in Providing Timely and Tailored Writing Feedback. Elementary school teachers pointed out their unique challenges as public school teachers. Due to the large

number of students in a single class and with only one teacher to deal with the class, teachers pointed out the difficulties of providing immediate and helpful feedback that support students in writing. T9 emphasized,

"I think providing individual feedback is a really time-consuming thing. It is difficult to individualize education for all subjects."

One of the teachers (T1), a director who has specialized in teaching writing in the writing center at one of the California school districts for the past 30 years, stressed the importance of developing ideas and adding details. T1 stated,

"I think for me, it seems like the area where kids need the most support is actually generating ideas for writing and adding details. Students might give you a sentence or two and say I'm done. But if teachers or AI ask them to add more details, that could enhance their writing. Such as asking, 'Can you tell me more about this?'—we can encourage them to expand their writing. Students frequently find it challenging to elaborate on their own without such guidance."

Our findings suggest that teachers can leverage GAI to provide immediate feedback regarding students' writing progress from ideation, grammar checkers, and adding detail. For example, T4 highlighted,

"I would love for AI to be able to do this for my students. Could AI give high-quality feedback on the spot to student writing? So I would love for the AI assistant to say, oh, you only used the word pretty. Is there another way to explain it? Can you provide some examples of your opinions? Can you explain more about your character?."

In this regard, interacting with GAI systems (i.e., LLM chatbot, TTI generators) helps students expand their ideas by enabling rapid prototypes that broaden their options. As one of the students (S10) stated,

"Since AI provides many options, I can pick the one I like best. I think it is good for me to come up with more ideas because AI has given me suggestions I never thought of, even when I get unexpected results, which actually makes me think of better ideas. Thanks to AI, I think the process went much faster."

According to our findings, using GAI systems would benefit teachers and students. Teachers can reduce the effort they need to provide individual attention to students, and students will be able to receive feedback on their story creation through GAI systems conversation.

4.2.3 Culturally Inclusive Pedagogy: GAI's Capabilities for Culturally Relevant Literacy Feedback. The other aspect of using GAI for personalized learning is to provide culturally relevant feedback and ideas [80]. Teachers and parents were particularly intrigued about the possibility of translating languages and providing examples of different cultures with GAI systems. Teachers intend to utilize GAI systems to generate culturally tailored examples they might not be familiar with during lesson planning. T15 stated,

"If I'm giving an assignment, and I'm trying to give examples, I only know the examples I know. And I have my cultural bias, I have my background, my limited experience. But if I get to ChatGPT to generate more examples of active and passive voice, it's gonna save a lot of time. And again, I can incorporate things from different interest levels, cultures, and vocabulary levels."

In our workshop, one of the parents shared that she used ChatGPT to generate word problems for her child's home language learning, which was Japanese. As a parent of an immigrant child, she wanted her daughter to remain fluent in her mother language. Also, parents who immigrated from Asia mentioned that they are willing to use GAI systems to create culturally salient fable stories that fit their children's interests. According to P04,

"So maybe parents will ask to know something about some traditional stories about their own culture, but they don't have the actual book or the graphic reference, like some Asian stories in China, about maybe a dragon or something, maybe parents will ask, do you know how to draw a Chinese dragon? And AI will say the Chinese dragon looked like a really long snake with some hair on the head. Also, they speak different languages. I think language translation will also be another activity, like my kids having Korean friends from Korea. So they want to share some Korean as well."

The finding indicates that the potential advantages of using GAI systems are to help teachers create lessons using culturally relevant materials, such as songs, videos, and images (i.e., traditional stories by countries' traditional holidays). By doing so, teachers can create a more culturally inclusive classroom and foster cross-cultural understanding. Additionally, parents, especially those from multicultural families, could bridge the communication gap between each other and encourage a sense of belonging and a strong family relationship through a better understanding of each other's cultural values.

4.3 Navigating the Gray Areas: Challenges and Constraints of GAI in Literacy Education

Our research indicates that GAI systems in academic settings may pose challenges related to academic integrity, such as issues of authorship, authenticity, and originality. Additionally, there are concerns about how these systems may impact student agency and autonomy in writing processes. A notable risk is the potential for GAI systems to generate biased or inaccurate content stemming from their inherent randomness and uncertainty.

4.3.1 Ethical Quandaries, and Accountability in GAI-LLM Writing systems. Nine teachers (9/16) expressed concerns about introducing GAI systems to their students due to the possibility of affecting the originality of their students' work. To teachers, AI-generated work can be a problem for kids to misrepresent themselves. As T06 stressed,

"So it's like, if you are using this as a tool, you're taking this work from somewhere, right? Make it your own and claim it your own. I think that the problem is that you took AI, and you didn't give AI the credit. If you're going to use AI, then that's who should be credited for the work of GPT because there's almost a moral issue for me, looking at Chat GPT. And thinking about where that information comes from."

By extension, teachers are anxious about GAI systems because if students use ChatGPT to generate their own work, it could undermine students' reasoning. For instance, T16 emphasized,

"I mean, teachers are particularly anxious about maintaining the quality of writing and are worried about students' work ethic and creativity. Additionally, there's a significant concern regarding plagiarism and cheating."

Some teachers, in response, suggested using ChatGPT rather than generating text as an output for students' writing, asking

students questions to promote the students' thought processes. As T12 mentioned,

"Here's one thing is, instead of writing the whole next part. It asked me, you know, like, choose your own adventure? Do you want it to be this kind of problem or that kind of problem? What comes next?"

As far as implementation plans were concerned, seven teachers (7/16) emphasized the need for the school districts and educators to establish a new framework for adapting GAI systems to students' learning, with (6/12) teachers also pointing out the necessity to establish different assessment methods.

The findings demonstrate the importance of designing the GAI systems to promote students' reasoning by providing students with the opportunity to use their own critical thinking skills and creative solutions. Additionally, educators must develop a new means of assessing and evaluating students' writing projects. For instance, teachers can focus on students' learning processes rather than their outcomes, asking their thoughts and opinions instead of asking them to write a certain number of words or paragraphs. This can help to identify areas of strength and weakness in the students' writing and help them to develop their writing skills.

4.3.2 The Agency Dilemma: Unpacking Student Agency in the Complex Role of GAI in Student Literacy. There has been difficulty determining the level of agency students have over their writing outputs when using GAI, particularly when it comes to disambiguating how much students write (i.e., the ideas, the sentence, the paragraph, the word choice) versus what GAI suggests and generates. From the writing workshop, we observed that many students (8/12) just copied and pasted directly from GAI-generated outputs into a Google Doc (i.e., "I'm done, I like the story, so why should I change it?"), raising the question of how to design the system to promote the craft of writing, such as idea generation, voice and style, audience awareness, revising, and more. Perceiving cutting and pasting as a reflection of a lack of agency by their children, parents were skeptical about the impact GAI would have on their children. Most parents (9/12) pointed out the importance of establishing fundamental knowledge first (i.e., comprehension and critical thinking skills) before introducing such automated systems as ChatGPT. As P07 mentioned,

"How do my kids learn if AI generates everything for them? And do they know enough about the content of what they're asking the AI? I think learning is trial and error by doing things by themselves, and kids need to have the foundation to be able to build upon to access that new AI."

Other parents consider that AI system access should determined by age-appropriate standards, as P11 stated,

"I think the current version is definitely not for kids age 8 or 9, it's too open-ended, my kid is too young and it's more important to learn foundation knowledge first, I think that there is learning that has to happen with that."

In accordance with the previous section, one of the key questions raised by adult participants was aspects of student autonomy (their ownership and agency over their writing project). The issue raises the challenge of designing child-AI interaction so that children can control their own learning processes, not just be led by AI. Hence, it is essential to develop AI-driven systems that respect children's autonomy, provide them with appropriate guidance and support, and ensure that the systems are suitable for children's age groups.

4.3.3 Erratic Outputs: Limitations and Concerns in Deploying GAI for Literacy Education. Like any generative AI chatbot or voice assistant–such as Siri, Alexa, or Microsoft's ill-fated Tay [107]–some individuals intentionally try to corrupt or manipulate GAI-produced responses, particularly in online settings. This behavior can take various forms, including providing chatbots with inappropriate or harmful content to elicit inappropriate responses, pushing the boundaries of what the chatbot can understand or respond to by inputting nonsensical or unusual queries to see how the chatbot reacts, or intentionally feeding chatbots with biased or false information to manipulate the responses and promote a particular agenda, ideology, or misinformation. The potential for such student-AI interactions was not lost on our teachers. T12 stressed,

"I can imagine there will be kids who want to test the limits and get the chatbot to say inappropriate things back to them. So, I mean, there's that part of it."

For instance, in our workshop, we observed students generating images around inappropriate political scenes (i.e., a Hitler statue), pointing to the need for developers of educational chatbot systems to implement safeguards and moderation mechanisms to minimize the impact of such intentional abuse. These safeguards may include content filtering, moderation of user inputs, and continuous improvement of the chatbot's response mechanisms to detect and handle inappropriate or harmful content [36, 81, 92]. Less malicious but still disruptive are instances where a GAI system produces surreal or nonsensical responses to user prompts. GAI hallucination, also known as AI-generated hallucination or AI-induced hallucination, refers to a phenomenon where generative models produce content that may resemble hallucinations in humans, including images, text, or other sensory data that are typically unintended and often nonsensical (i.e., a dead animal without a head). AI hallucination occurs when a machine learning model generates content that doesn't align with the intended output [13, 18]. It can result from the model's overfitting to its training data, exposure to unusual or biased data, or other factors that cause the model to produce strange or distorted outputs. A student (S02) pointed out an unexpected result had been generated from the GAI systems and stated,

"If I do it without artificial intelligence, I can do it with my hands exactly as I thought, but if I use artificial intelligence, I think it can be seen as a disadvantage in that it is expressed slightly differently than my intention."

In instances where GAI-produced content is inaccurate but seemingly plausible, parents (10/12) argued it is important to consider whether or not students know AI-provided information is accurate. Several parents cited the need for educational AI deployments to be prefaced with fundamental education to develop critical thinking, comprehension, and problem-solving skills so their child can critically analyze and scrutinize information:

"And do they know enough about the content of what they're asking the AI? How do we know if kids ask the right question, and how do we know if the information provided by AI is correct or not for students? I think kids first learn through credited resources and develop that fundamental knowledge, at least by middle school."

Based on our findings, we identified several challenges with current GAI systems, including the originality of students' writing projects (academic integrity), the agency of students in writing processes (learning), and the generation of misinformation due to the randomness of the GAI systems. These challenges are not distinctive from one another; rather, they are interconnected and need to be addressed collectively. In section 6, we discuss design implications that address the challenges mentioned above.

5 DISCUSSION

From the study, we examined the potential advantages and challenges of using GAI systems for literacy education in K-6 settings from multiple stakeholders' perspectives. We discovered how each stakeholder's views differ: for teachers, generative AI systems are a new type of digital citizenship development; for parents, these GAI systems are another type of toys or games; for students, these are smart, helpful companions.

In our discussion, we delve into the complexities of integrating cutting-edge educational technologies into learning settings, scrutinizing their impact on the design of GAI learning systems. Additionally, we outline three key design considerations essential for developing effective GAI-based educational applications.

5.1 Unpacking the Complexity of Technology Integration in Education

Despite substantial investments in educational technology, there is often a notable gap between the anticipated and actual usage of these tools in classroom environments [27]. Teachers' varying levels of comfort and proficiency with technology significantly influence its application in teaching. Resource limitations also pose significant challenges, with issues like inadequate training, support, and access to current and functional technology impeding effective utilization [27]. Reich (2020) underscores the importance of addressing the broader social, cultural, and pedagogical complexities in education, which he deems more crucial than mere technological advancement [87].

The recent LLMs have brought breakthroughs of open-ended conversational systems, which perform open-domain dialog with any topics [51] and it offers the capability to be fine-tuned [82], enhancing its performance to align with specific domains and instructional objectives [112, 129]. Unlike traditional MOOC platforms, which rely on human-guided instructions, educators can now train the LLM with specialized datasets and employ prompt engineering techniques [114] to enable AI to construct instructional content autonomously. Furthermore, students' educational behavior data, which includes their challenges and areas of proficiency, can be fed back into the LLM for evaluation. This allows for algorithmically-guided decisions about where to begin instruction based on each student's capabilities. Eventually, educational systems will likely converge three distinct approaches within an integrated system—combining direct instruction, algorithm-guided learning, and AI facilitation. This system will not only instruct and guide but also foster open-ended exploration and collaboration between students and AI agents. Therefore, it is important to examine the possibility that these GAI systems can be integrated with new pedagogical approaches.

Consequently, new breakthrough systems like GPTs [79] will require thorough evaluation in terms of safety, effectiveness, and their ability to foster trust and community integration before they gradually become embedded in societal norms. Organizations such as the Institute of Education Sciences (IES) and Digital Promise, among others, are beginning to form communities of educators to explore the possibilities these systems offer and to critically examine their applicability for teaching and learning [7, 56]. Consequently, it is anticipated that these technologies will be integrated into educational systems gradually rather than affecting a radical transformation in teaching and learning methodologies immediately.

5.1.1 Double-Edged Sword of GAI in Education. From the study, we found educators were drawn to use the GAI systems for instruction and in the way that creating lesson plans (e.g., pre-, during, and post-instruction) can be made easier using AI-scaffolded content creation. Meanwhile, students found they could leverage the systems to receive individualized and timely feedback. At the same time, parents pointed out the GAI systems' capabilities to facilitate interest-driven learning, particularly about culturally relevant approaches in writing projects [9, 83].

The use of GAI in educational settings presents a complex blend of benefits and drawbacks, which are not mutually exclusive but rather exist simultaneously, reflecting a double-edged nature. GAI facilitates open dialogue and free-form conversation, enabling the exploration of culturally diverse topics and translation capabilities. This openness enriches the educational experience by fostering a broader understanding of various cultures and languages. On the other hand, the same openness of GAI systems can lead to potential challenges, including the development of biased perspectives and the generation of inaccurate or 'hallucinated' results [32]. Such issues underscore the critical need for careful moderation and strategic oversight, such as the implementation of customized models [30] (e.g., incorporating more diverse races into the image data set to train TTL) so that the system does not generate a particular ethnicity or race. Such precautions are crucial to harness the benefits of GAI while minimizing its risks for educational settings.

5.2 Recommendations for System Designers and Developers

As part of this discussion, we propose the design considerations of GAI-powered writing platforms to inform the designing of safe and accessible GAI systems for elementary school settings. To capitalize on the perceived benefits of educational uses of GAI while mitigating the concerns from our stakeholder groups, educational GAI platforms should: 1) provide guardrails to protect students' authorship issues in GAI-powered writing, 2) afford appropriate role allocation to AI and students, and 3) support customizable teacher-in-the-loop systems to enhance the trustworthiness and content-focus of GAI systems.

5.2.1 Navigating the Complexity of Authorship and Ownership in AI-Assisted Writing Systems. Our findings highlighted that teachers are concerned about their students' authorship and integrity of their writing output, particularly when GAI generates the majority of the content for students [25, 28, 64]. Even though studies have examined GAI-LLM-powered writing systems, such as Gero et al.

[42], Lee et al. [63], and Yuan et al. [126], focus on investigating language models' capacity rather than users' capabilities and their perspectives (including those with different cognitive levels, abilities, and ages). Furthermore, there is a lack of studies focusing on educational settings for K-6, which aim to mitigate specific problems they face (i.e., authorship, plagiarism, assessment) [104]. Gero et al. [42], and Lee et al. [63] have identified that there is no one-size-fits-all solution when it comes to users' sense of ownership and authorship over AI-assisted writing processes due to uncertainty over authorship of language model-generated texts itself. Consequently, there is a need for further research into writers' ownership, authorship, and plagiarism, in addition to developing new methods for assessing and measuring writers' progress [29, 58, 59].

To better understand what guardrails and guidelines need to be implemented into the development of GAI-LLM-powered co-writing systems for K-6 students, future research on students' capacity, especially on measuring learning processes and assessment of the writing (e.g., how they interact with GAI-LLM like ChatGPT), would be beneficial.

To navigate the authorship and ownership of AI-assisted writing systems like ChatGPT, we propose building a system based on the LLM that facilitates cloud-based infrastructure. The database stores students' utterances in separation from AI-generated texts. To differentiate between student-generated content and machine-generated text, the platform will employ text-similarity analysis [57]. This method allows educators to compare student writing with AI output, offering insights into the extent of AI reliance on student work.

5.2.2 Enhancing Student Agency through Role Allocation in GAI Systems Design. We observed that when students encountered open-ended GAI systems' interfaces (i.e., ChatGPT and Stable Diffusion) without context, they had difficulty writing prompts in a way that produced appropriate results. Hence, we argue for designing GAI-LLM co-writing platforms that mimic natural conversation, providing students with concrete context at the beginning of their interaction and offering options for choosing topics of choice and characters to support child-centered and interest-driven learning experiences [34, 83].

According to the workshop with students and teacher interviews, promoting students' agency as writers is essential [62], especially for enhancing learning experiences. As a result, students should be given opportunities to participate in writing projects and promote independent writing actively. This can be accomplished by allowing students to customize and edit their own writing. To facilitate safer and more efficient GAI systems in education without compromising their integrity, system developers and Edtech designers need to establish a division of tasks, setting up boundaries of roles between the AI agent for educators and students. By designing an AI agent persona and curating Child-AI conversations, this can be achieved by encouraging idea generation, adding story detail, and elaborating from the perspective of students. AI agents should be designed to help students think critically and creatively and to encourage them to ask questions through conversation [6, 123]. For instance, system developers allocate AI's persona as a coach or/peer rather than an assistant– that means rather than having AI generate writing on students' behalf, designing AI agents that encourage students to write their own creative ideas, giving students control over the writing process. Nguyen. [76] discusses the benefits of designing prompts that enable chatbots to foster systemic thinking (such as idea generation and questioning). Specifically, Nguyen. [76] examined textual conversational agents' (chatbot) role design (personas) and its impact on students' system thinking process in group discussions. The findings suggested more transactive exchanges with less knowledgeable peer agents (versus interacting with expert agents) as students felt more social and engaging. This finding suggests that designing an age-appropriate agent role/ persona can impact conceptual understanding, enhancing learning outcomes. The current capacity of prompting LLMs offers possibilities to optimize the free-form LLM-based chatbot dialogues for that purpose.

5.2.3 Balancing Flexibility and Control GAI-LLM Systems for Educator and Parent Oversight. Our findings indicated that teachers and parents expressed concern about students' interaction with misinformation and biased content due to the system's randomness. To mitigate the uncertainty associated with GAI LLM systems, it is essential to design a system that balances flexibility and control with adults-in-the-loop systems [49, 72, 128]. Yuan et al. [126] examined some of the methods that oversee the writing processes by providing suggestion options for users and offering prompt design features from the back end. However, deciding and accepting the suggestions and writing prompts could be challenging for a certain age group and intellectual level or English proficiency [69].

Hence, we propose designing an 'educators' view' that allows educators and/or parents to easily 'prompt' and curate GAI-based chatbots' conversation to facilitate a secure mode of student-AI interaction for writing. For example, the new systems will allow educators to prompt GAI systems to carry on their lessons, similar to the current tool that designs a chatbot with flow-based interfaces, such as Voiceflow [8]. Our suggestion is to develop flow-based interfaces [37] (or block-based interfaces [12]) for educators, where each node or block can translate into a prompt, which will create dialogue as teachers intend, continue writing project instructions, and construct conversation for students. By doing so, the system will provide educators control over a certain level of uncertainty the current GAI-LLM-based chatbot might have and provide open-ended flexibility, with low floors and high ceilings [89].

The majority of teachers (12/16) we interviewed expressed difficulty adapting to new tools and AI applications (due to their heavy workload). Therefore, interfaces should be as simple (and easy to use) as teachers already know. To design the system, we recommend actively collaborating with teachers, co-designing the processes and interfaces through multiple steps of studies starting with need-findings and card-sorting [95] to understand their unique languages and mental model to create an appropriate conceptual model that aligns with educators' goals [65]. With that series of user tests and gathering feedback from teachers and students, it is possible to refine the system and optimize its functionalities for educational purposes.

5.3 Directions for the future work

For future research directions aimed at broadening the scope and generalizability of our findings, we advocate for an expanded investigation into GAI utilization. This should involve a comprehensive

analysis of system logs and behavioral data within GAI platforms. This includes leveraging GAI platforms for collecting back-end educational data to analyze students' learning progress such as their reliance on AI, writing quality, and the nature of AI-student interactions. By engaging a wider participant base and adopting a longitudinal study approach, we can deepen our understanding of how GAI tools influence user interactions, experiences, and learning outcomes over time.

To promote accelerated learning through GAI-powered learning tools, further research could also include A/B testing, using multidimensional metrics to evaluate student writing. These metrics include *Production*: the amount of writing users generate over time and per session within the system, *Narrativity*: the extent to which a text tells a story with characters, events, places, and things, *Syntactic Complexity*: the complexity of the text's syntactic structure, *Vocabulary*: sophistication and concreteness of students' word choice, *Grammatical Correctness*: the extent to which students' texts adhere to grammar norms [44, 45, 73]. By integrating these AI and database systems, designers and researchers will be better equipped to understand the details of student interaction with AI in writing, aiding in the development of more effective educational tools. This approach enriches insights into AI's educational applications and also sets a foundation for future studies focused on the nuanced dynamics of AI-assisted learning.

6 CONCLUSION

In this paper, we explored the stakeholders in education's perceptions and opinions regarding the advantages and limitations of leveraging GAI systems in literacy education for elementary school students. Through qualitative studies, conducting workshops and interviews with teachers, parents, and students of 40 total participants, we found that the GAI systems can be used to generate adaptive lesson plan materials such as mentor text for teachers for them to tailor according to each student's needs and skill level (through scaffolding and their interests). The GAI system affords culturally relevant and timely feedback that broadens ideation for writing projects. We also discovered the limitations of the systems in determining the authenticity of students' writing projects, difficulties determining students' agency over their writing outcomes, and concerns regarding the safety and accuracy of the content. Based on the findings, we provide implications for future studies to navigate authorship and ownership of AI-assisted writing projects that students produce. We also drew design suggestions to mitigate the concerns regarding the safety and accuracy of content. First, we recommend promoting student agency through role allocation over AI and humans, allowing more room for students to customize and edit their own writing. Second, we propose facilitating teacher-in-the-loop systems where educators and parents can control the lessons by prompting AI to carry on their lessons based on their design. Our study highlights an opportunity to foster collaboration between researchers in the HCI, Education, GAI, and NLP communities to design a GAI-powered platform for literacy education.

ACKNOWLEDGMENTS

We would like to express our gratitude to the participants for their invaluable contributions to our study. Special thanks to Daniel Epstein, Elena Agapie, and Aure Schrock for their insightful feedback and discussions. This research was made possible by the generous support of the VITAL Prize Challenge award.

REFERENCES

[1] Forest Agostinelli, Mihir Mavalankar, Vedant Khandelwal, Hengtao Tang, Dezhi Wu, Barnett Berry, Biplav Srivastava, Amit Sheth, and Matthew Irvin. 2021. Designing Children's New Learning Partner: Collaborative Artificial Intelligence for Learning to Solve the Rubik's Cube. In *Proceedings of the 20th Annual ACM Interaction Design and Children Conference* (Athens, Greece) *(IDC '21)*. Association for Computing Machinery, New York, NY, USA, 610–614. https://doi.org/10.1145/3459990.3465175

[2] Norita Ahmad, San Murugesan, and Nir Kshetri. 2023. Generative Artificial Intelligence and the Education Sector. *Computer* 56, 6 (June 2023), 72–76. https://doi.org/10.1109/MC.2023.3263576 Conference Name: Computer.

[3] MagicSchool ai. 2023. AI for teachers - lesson planning and more! https://www.magicschool.ai/

[4] Open AI. 2023. OpenAI API. https://platform.openai.com

[5] Speak ai. 2023. The speaking app that actually talks. https://www.speak.com/

[6] Mehdi Alaimi, Edith Law, Kevin Daniel Pantasdo, Pierre-Yves Oudeyer, and Hélène Sauzeon. 2020. Pedagogical Agents for Fostering Question-Asking Skills in Children. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376776

[7] Colin Angevine, Karen Cator, Babe Liberman, Kim Smith, and Viki Young. 2019. *Designing a Process for Inclusive Innovation: A Radical Commitment to Equity*. Technical Report. Digital Promise. https://doi.org/10.51388/20.500.12265/86

[8] Voiceflow application. 2023. Build amazing conversational assistants. https://www.voiceflow.com/

[9] Brittany Aronson and Judson Laughter. 2016. The Theory and Practice of Culturally Relevant Education: A Synthesis of Research Across Content Areas. *Review of Educational Research* 86, 1 (March 2016), 163–206. https://doi.org/10.3102/0034654315582066

[10] David Baidoo-Anu and Leticia Owusu Ansah. 2023. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. https://doi.org/10.2139/ssrn.4337484

[11] Erin Beneteau, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication Breakdowns Between Families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300473

[12] Google Blockly. 2023), publisher=Google. https://developers.google.com/blockly

[13] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

[14] Once Upon a Bot. 2023. Once upon a bot. https://onceuponabot.com/

[15] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. https://doi.org/10.48550/arXiv.2005.14165 arXiv:2005.14165 [cs].

[16] Bengisu Cagiltay, Bilge Mutlu, and Joseph E Michaelis. 2023. "My Unconditional Homework Buddy:" Exploring Children's Preferences for a Homework Companion Robot. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) *(IDC '23)*. Association for Computing Machinery, New York, NY, USA, 375–387. https://doi.org/10.1145/3585088.3589388

[17] Miguel A Cardona, Roberto J Rodríguez, and Kristina Ishmael. 2023. Artificial Intelligence and the Future of Teaching and Learning. (2023).

[18] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. 2021. Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)*. 2633–2650.

[19] ChatGPT. 2023. Chatgpt. https://openai.com/chatgpt

[20] ChatGPT. 2023. Introducing ChatGPT. https://openai.com/blog/chatgpt

[21] Lijia Chen, Pingping Chen, and Zhijian Lin. 2020. Artificial Intelligence in Education: A Review. *IEEE Access* 8 (2020), 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510 Conference Name: IEEE Access.

[22] Thomas K. F. Chiu, Qi Xia, Xinyan Zhou, Ching Sing Chai, and Miaoting Cheng. 2023. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence* 4 (Jan. 2023), 100118. https://doi.org/10.1016/j.caeai.2022.100118

[23] Irene-Angelica Chounta, Emanuele Bardone, Aet Raudsep, and Margus Pedaste. 2022. Exploring teachers' perceptions of Artificial Intelligence as a tool to support their practice in Estonian K-12 education. *International Journal of Artificial Intelligence in Education* 32, 3 (2022), 725–755. https://doi.org/10.1007/s40593-021-00243-5

[24] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 329–340. https://doi.org/10.1145/3172944.3172983

[25] Debby RE Cotton, Peter A Cotton, and J Reuben Shipway. 2023. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International* (2023), 1–12.

[26] Mihaly Csikszentmihalyi. 2009. *Flow: The Psychology of Optimal Experience*. Harper Collins. Google-Books-ID: QVjPsd1UukEC.

[27] Larry Cuban. 2001. *Oversold and underused: Computers in the classroom*. Harvard university press.

[28] Geoffrey M Currie. 2023. Academic integrity and artificial intelligence: is ChatGPT hype, hero or heresy?. In *Seminars in Nuclear Medicine*. Elsevier.

[29] Nassim Dehouche. 2021. Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3). *Ethics in Science and Environmental Politics* 21 (2021), 17–23.

[30] Paul DelSignore. 2023. Train your own image model using Leonardo AI. https://medium.com/the-generator/train-your-own-image-model-using-leonardo-ai-7a7e92d1a387

[31] Stable Diffusion. 2023. Stable diffusion online. https://stablediffusionweb.com/

[32] Carmen Drahl. 2023. Ai was asked to create images of black African docs treating White Kids. how'd it go? https://www.npr.org/sections/goatsandsoda/2023/10/06/1201840678/ai-was-asked-to-create-images-of-black-african-docs-treating-white-kids-howd-it-

[33] Stefania Druga, Fee Lia Christoph, and Amy J Ko. 2022. Family as a Third Space for AI Literacies: How do children and parents learn about AI together?. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[34] Daniel C Edelson and Diana M Joseph. 2012. The interest-driven learning design framework: motivating learning through usefulness. In *Embracing Diversity in the Learning Sciences*. Routledge, 166–173.

[35] Maha Elgarf, Sahba Zojaji, Gabriel Skantze, and Christopher Peters. 2022. CreativeBot: a Creative Storyteller robot to stimulate creativity in children. In *Proceedings of the 2022 International Conference on Multimodal Interaction (ICMI '22)*. Association for Computing Machinery, New York, NY, USA, 540–548. https://doi.org/10.1145/3536221.3556578

[36] Joel E Fischer. 2023. Generative AI Considered Harmful. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 7, 5 pages. https://doi.org/10.1145/3571884.3603756

[37] Vue Flow. 2023. vueflow: the customizable Vue3 Flowchart Library. https://vueflow.dev/

[38] Giorgio Franceschelli and Mirco Musolesi. 2023. On the creativity of large language models. *arXiv preprint arXiv:2304.00008* (2023).

[39] Radhika Garg, Hua Cui, Spencer Seligson, Bo Zhang, Martin Porcheron, Leigh Clark, Benjamin R. Cowan, and Erin Beneteau. 2022. The Last Decade of HCI Research on Children and Voice-Based Conversational Agents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 149, 19 pages. https://doi.org/10.1145/3491102.3502016

[40] 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* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376631

[41] 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.

[42] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing Using Language Models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1002–1019. https://doi.org/10.1145/3532106.3533533

[43] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 245, 15 pages. https://doi.org/10.1145/3544548.3580782

[44] Arthur C. Graesser, Danielle S. McNamara, Zhiqang Cai, Mark Conley, Haiying Li, and James Pennebaker. 2014. COH-metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal* 115, 2 (2014), 210–229. https://doi.org/10.1086/678293

[45] Arthur C. Graesser, Danielle S. McNamara, and Jonna M. Kulikowich. 2011. Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher* 40, 5 (June 2011), 223–234. https://doi.org/10.3102/0013189X11413260 Publisher: American Educational Research Association.

[46] Steve Graham, Alisha Bollinger, Carol Booth Olson, Catherine D'Aoust, Charles MacArthur, Deborah McCutchen, and Natalie Olinghouse. 2012. Teaching Elementary School Students to Be Effective Writers: A Practice Guide. NCEE 2012-4058. *What Works Clearinghouse* (2012).

[47] Steve Graham, Karen R Harris, Sharlene A Kiuhara, and Evan J Fishman. 2017. The relationship among strategic writing behavior, writing motivation, and writing performance with young, developing writers. *The Elementary School Journal* 118, 1 (2017), 82–104.

[48] Steve Graham and Dolores Perin. 2007. Writing next-effective strategies to improve writing of adolescents in middle and high schools.

[49] Adit Gupta and Christopher J MacLellan. 2021. Designing Teachable Systems for Intelligent Tutor Authoring. (2021).

[50] Ariel Han and Zhenyao Cai. 2023. Design Implications of Generative AI Systems for Visual Storytelling for Young Learners. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) *(IDC '23)*. Association for Computing Machinery, New York, NY, USA, 470–474. https://doi.org/10.1145/3585088.3593867

[51] Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in Building Intelligent Open-Domain Dialog Systems. *ACM Trans. Inf. Syst.* 38, 3, Article 21 (apr 2020), 32 pages. https://doi.org/10.1145/3383123

[52] Bernd Huber, Daniel McDuff, Chris Brockett, Michel Galley, and Bill Dolan. 2018. Emotional Dialogue Generation using Image-Grounded Language Models. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173851

[53] Seth Hunter, Jeevan Kalanithi, and David Merrill. 2010. Make a Riddle and TeleStory: designing children's applications for the siftables platform. In *Proceedings of the 9th International Conference on Interaction Design and Children (IDC '10)*. Association for Computing Machinery, New York, NY, USA, 206–209. https://doi.org/10.1145/1810543.1810572

[54] Anealka Aziz Hussin. 2018. Education 4.0 made simple: Ideas for teaching. *International Journal of Education and Literacy Studies* 6, 3 (2018), 92–98.

[55] Gwo-Jen Hwang, Haoran Xie, Benjamin W Wah, and Dragan Gašević. 2020. Vision, challenges, roles and research issues of Artificial Intelligence in Education. , 100001 pages.

[56] IES. 2023. Institute of Education Sciences. https://ies.ed.gov/ncer/projects/program.asp?ProgID=13

[57] Aminul Islam and Diana Inkpen. 2008. Semantic Text Similarity Using Corpus-Based Word Similarity and String Similarity. *ACM Trans. Knowl. Discov. Data* 2, 2, Article 10 (jul 2008), 25 pages. https://doi.org/10.1145/1376815.1376819

[58] Mohammad Khalil and Erkan Er. 2023. Will ChatGPT get you caught? Rethinking of plagiarism detection. *arXiv preprint arXiv:2302.04335* (2023).

[59] Michael R King and ChatGPT. 2023. A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and Molecular Bioengineering* 16, 1 (2023), 1–2.

[60] Alyson Klein. 2023. 180 Degree Turn: NYC District Goes From Banning ChatGPT to Exploring AI's Potential. *Education Week* (Oct. 2023). https://www.edweek.org/technology/180-degree-turn-nyc-schools-goes-from-banning-chatgpt-to-exploring-ais-potential/2023/10

[61] DiCerbo Kristen. 2023. Khan Academy explores the potential for GPT-4 in a limited pilot program. https://openai.com/customer-stories/khan-academy

[62] Natalia Kucirkova. 2019. Children's agency by design: Design parameters for personalization in story-making apps. *International Journal of Child-Computer Interaction* 21 (2019), 112–120. https://doi.org/10.1016/j.ijcci.2019.06.003,doi={10.1016/j.ijcci.2019.06.003},publisher={Elsevier}

[63] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3491102.3502030

[64] Weng Marc Lim, Asanka Gunasekara, Jessica Leigh Pallant, Jason Ian Pallant, and Ekaterina Pechenkina. 2023. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education* 21, 2 (2023), 100790. https://doi.org/10.1016/j.ijme.2023.100790

[65] Phoebe Lin and Jessica Van Brummelen. 2021. Engaging Teachers to Co-Design Integrated AI Curriculum for K-12 Classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3411764.3445377

[66] Maria Lindh and Jan Nolin. 2016. Information we collect: Surveillance and privacy in the implementation of google apps for education. *European Educational Research Journal* 15, 6 (2016), 644–663. https://doi.org/10.1177/1474904116654917

[67] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3491102.3501825

[68] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3526113.3545621

[69] Becton Loveless. 2023. Learning to Read to Read to Learn. Myth or Reality? https://www.educationcorner.com/learning-to-read-to-read-to-learn.html

[70] Jacqueline Low. 2019. A pragmatic definition of the concept of theoretical saturation. *Sociological Focus* 52, 2 (Jan 2019), 131–139. https://doi.org/10.1080/00380237.2018.1544514

[71] Wander ly. 2023. Wanderly: Magical adventures for curious children. https://app.wander.ly/

[72] Christopher J MacLellan, Rob Sheline, and Erik Harpstead. 2019. A Human-Centered Approach to Designing Teachable Systems. (2019).

[73] Danielle S. McNamara, Scott A. Crossley, and Philip M. McCarthy. 2010. Linguistic Features of Writing Quality. *Written Communication* 27, 1 (Jan. 2010), 57–86. https://doi.org/10.1177/0741088309351547 Publisher: SAGE Publications Inc.

[74] Midjourney. 2023. Midjourney. https://www.midjourney.com/app/

[75] Michael Muller, Lydia B Chilton, Anna Kantosalo, Charles Patrick Martin, and Greg Walsh. 2022. GenAICHI: Generative AI and HCI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3491101.3503719

[76] Ha Nguyen. 2023. Role design considerations of conversational agents to facilitate discussion and systems thinking. *Computers & Education* 192 (Jan. 2023), 104661. https://doi.org/10.1016/j.compedu.2022.104661

[77] Eric Nichols, Leo Gao, and Randy Gomez. 2020. Collaborative storytelling with large-scale neural language models. In *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*. 1–10.

[78] OpenAI. 2022. DALL·E 2. https://openai.com/dall-e-2/

[79] OpenAI. 2023. Introducing gpts. https://openai.com/blog/introducing-gpts

[80] Django Paris and H Samy Alim. 2017. *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world*. Teachers College Press.

[81] Pat Pataranutaporn, Valdemar Danry, Lancelot Blanchard, Lavanay Thakral, Naoki Ohsugi, Pattie Maes, and Misha Sra. 2023. Living Memories: AI-Generated Characters as Digital Mementos. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 889–901. https://doi.org/10.1145/3581641.3584065

[82] Andrew Peng, Michael Wu, John Allard, Logan Kilpatrick, and Steven Heidel. 2023. https://openai.com/blog/gpt-3-5-turbo-fine-tuning-and-api-updates

[83] Kylie Peppler. 2013. New Opportunities for Interest-Driven Arts Learning in a Digital Age. *Wallace Foundation* (2013).

[84] Igor Pesek, Novica Nosović, and Marjan Krašna. 2022. The Role of AI in the Education and for the Education. In *2022 11th Mediterranean Conference on Embedded Computing (MECO)*. IEEE, 1–4.

[85] Sara Polak, Gianluca Schiavo, and Massimo Zancanaro. 2022. Teachers' Perspective on Artificial Intelligence Education: An Initial Investigation. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 431, 7 pages. https://doi.org/10.1145/3491101.3519866

[86] Marc Prensky. 2001. Digital natives, digital immigrants part 2: Do they really think differently? *On the horizon* 9, 6 (2001), 1–6.

[87] Justin Reich. 2020. *Failure to disrupt: Why technology alone can't transform education*. Harvard University Press.

[88] Léonie J Rennie and Tina Jarvis. 1995. Three approaches to measuring children's perceptions about technology. *International Journal of Science Education* 17, 6 (1995), 755–774.

[89] Mitchel Resnick, Brad Myers, Kumiyo Nakakoji, Ben Shneiderman, Randy Pausch, and Mike Eisenberg. 2005. Design Principles for Tools to Support Creative Thinking. *Report of Workshop on Creativity Support Tools* 20 (Jan. 2005).

[90] Kalhan Rosenblatt. 2023. CHATGPT banned from New York City public schools' devices and Networks. https://www.nbcnews.com/tech/tech-news/new-york-city-public-schools-ban-chatgpt-devices-networks-rcna64446

[91] Kalhan Rosenblatt. 2023. New York City public schools remove chatgpt ban. https://www.nbcnews.com/tech/chatgpt-ban-dropped-new-york-city-public-schools-rcna85089

[92] Daniel Russell, Q. Vera Liao, Chinmay Kulkarni, Elena L. Glassman, and Nikolas Martelaro. 2023. Human-Computer Interaction and AI: What Practitioners Need to Know to Design and Build Effective AI System from a Human Perspective. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 545, 3 pages. https://doi.org/10.1145/3544549.3574170

[93] Johnny Saldaña. 2013. *The coding manual for qualitative researchers* (2nd ed ed.). SAGE, Los Angeles. OCLC: ocn796279115.

[94] Ismaila Temitayo Sanusi, Solomon Sunday Oyelere, Henriikka Vartiainen, Jarkko Suhonen, and Markku Tukiainen. 2023. A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies* 28, 5 (2023), 5967–5997.

[95] Katie Sherwin. 2018. Card sorting: Uncover users' mental models for better information architecture. https://www.nngroup.com/articles/card-sorting-definition/

[96] Rashi Shrivastava. 2022. Amazon Alexa wants to put your child to bed with generative ai storytelling. https://www.forbes.com/sites/rashishrivastava/2022/12/07/amazon-alexa-wants-to-put-your-child-to-bed-with-generative-ai-storytelling/?sh=29b2266eb5b9

[97] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2022. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Trans. Comput.-Hum. Interact.* (feb 2022). https://doi.org/10.1145/3511599 Just Accepted.

[98] Jacob Steiss, Tamara Tate, Steve Graham, Jazmin Cruz, Michael Hebert, Jiali Wang, Youngsun Moon, Waverly Tseng, et al. 2023. Comparing the Quality of Human and Chatgpt Feedback on Students' Writing. (2023).

[99] Jiahong Su, Davy Tsz Kit Ng, and Samuel Kai Wah Chu. 2023. Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence* 4 (2023), 100124.

[100] Ben Swanson, Kory Mathewson, Ben Pietrzak, Sherol Chen, and Monica Dinalescu. 2021. Story centaur: Large language model few shot learning as a creative writing tool. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. 244–256.

[101] Fati Tahiru. 2021. AI in education: A systematic literature review. *Journal of Cases on Information Technology (JCIT)* 23, 1 (2021), 1–20.

[102] David R Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation* 27, 2 (2006), 237–246.

[103] Atlas ti. 2023. The #1 software for qualitative data analysis. https://atlasti.com/

[104] Levent Uzun. 2023. ChatGPT and academic integrity concerns: Detecting artificial intelligence generated content. *Language Education and Technology* 3, 1 (2023).

[105] Jessica Van Brummelen, Maura Kelleher, Mingyan Claire Tian, and Nghi Nguyen. 2023. What Do Children and Parents Want and Perceive in Conversational Agents? Towards Transparent, Trustworthy, Democratized Agents. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) *(IDC '23)*. Association for Computing Machinery, New York, NY, USA, 187–197. https://doi.org/10.1145/3585088.3589353

[106] Maarten Van Mechelen, Rachel Charlotte Smith, Marie-Monique Schaper, Mariana Tamashiro, Karl-Emil Bilstrup, Mille Lunding, Marianne Graves Petersen, and Ole Sejer Iversen. 2023. Emerging Technologies in K–12 Education: A Future HCI Research Agenda. *ACM Trans. Comput.-Hum. Interact.* 30, 3, Article 47 (jun 2023), 40 pages. https://doi.org/10.1145/3569897

[107] James Vincent. 2016. Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a Day. https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

[108] Jessica Walrack, Barri Segal, and Susannah Snider. 2023. Where do I fall in the American Economic Class System? https://money.usnews.com/money/personal-finance/family-finance/articles/where-do-i-fall-in-the-american-economic-class-system

[109] Ge Wang, Kaiwen Sun, Ayça Atabey, Kruakae Pothong, Grace C Lin, Jun Zhao, and Jason Yip. 2023. Child-Centred AI Design: Definition, Operation, and Considerations. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6.

[110] Ge Wang, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2022. 'Don't make assumptions about me!': Understanding Children's Perception of Datafication Online. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.

[111] Ge Wang, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2023. 'Treat me as your friend, not a number in your database': Co-designing with Children to Cope with Datafication Online. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.

[112] Lu Wang, Munif Ishad Mujib, Jake Williams, George Demiris, and Jina Huh-Yoo. 2021. An evaluation of generative pre-training model-based therapy chatbot

for caregivers. *arXiv preprint arXiv:2107.13115* (2021).

[113] Suchen Wang, Yueqi Duan, Henghui Ding, Yap-Peng Tan, Kim-Hui Yap, and Junsong Yuan. 2022. Learning transferable human-object interaction detector with natural language supervision. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022). https://doi.org/10.1109/cvpr52688.2022.00101

[114] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. https://doi.org/10.48550/arXiv.2302.11382 arXiv:2302.11382 [cs].

[115] Randi Williams, Hae Won Park, and Cynthia Breazeal. 2019. A is for Artificial Intelligence: The Impact of Artificial Intelligence Activities on Young Children's Perceptions of Robots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300677

[116] Tongshuang Wu, Michael Terry, and Carrie J. Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts | Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. https://dl.acm.org/doi/abs/10.1145/3491102.3517582

[117] Ying Xu, Stacy Branham, Xinwei Deng, Penelope Collins, and Mark Warschauer. 2021. Are Current Voice Interfaces Designed to Support Children's Language Development?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Yokohama</city>, <country>Japan</country>, </conf-loc>) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 633, 12 pages. https://doi.org/10.1145/3411764.3445271

[118] Ying Xu, Stacy Branham, Xinwei Deng, Penelope Collins, and Mark Warschauer. 2021. Are Current Voice Interfaces Designed to Support Children's Language Development?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3411764.3445271

[119] Ying Xu, Stacy Branham, Xinwei Deng, Penelope Collins, and Mark Warschauer. 2021. Are Current Voice Interfaces Designed to Support Children's Language Development?. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–12.

[120] Ying Xu, Kunlei He, Valery Vigil, Santiago Ojeda-Ramirez, Xuechen Liu, Julian Levine, Kelsyann Cervera, and Mark Warschauer. 2023. "Rosita Reads With My Family": Developing A Bilingual Conversational Agent to Support Parent-Child Shared Reading. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) *(IDC '23)*. Association for Computing Machinery, New York, NY, USA, 160–172. https://doi.org/10.1145/3585088.3589354

[121] 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 (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3299035

[122] 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* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376416

[123] Özge Nilay Yalçin, Sebastien Lalle, and Cristina Conati. 2022. An Intelligent Pedagogical Agent to Foster Computational Thinking in Open-Ended Game Design Activities. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 633–645. https://doi.org/10.1145/3490099.3511124

[124] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, and Toby Jia-Jun Li. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. *ACM IUI Workshop* (2022).

[125] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376301

[126] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. https://doi.org/10.1145/3490099.3511105

[127] Chao Zhang, Cheng Yao, Jianhui Liu, Zili Zhou, Weilin Zhang, Lijuan Liu, Fangtian Ying, Yijun Zhao, and Guanyun Wang. 2021. StoryDrawer: A Co-Creative Agent Supporting Children's Storytelling through Collaborative Drawing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3411763.3451785

[128] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 2022. StoryBuddy: A Human-AI Collaborative Chatbot for Parent-Child Interactive Storytelling with Flexible Parental Involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3491102.3517479

[129] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2018. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (Apr. 2018). https://doi.org/10.1609/aaai.v32i1.11325