UNIVERSITY OF CALIFORNIA,
IRVINE

# StoryAI: designing, developing, and evaluating Generative AI-powered story-authoring platform for young learners

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Informatics

by

Ariel Han

Dissertation Committee:
Professor Kylie Peppler, Chair
Professor Katie Salen Tekinbaş
Professor Kurt Dean Squire

2024

# DEDICATION

To

my daughters, Jiyul and Sunyul.

You are the inspiration and purpose for me to work in the field of educational informatics.

You made me strong physically and mentally,

You made me want to be a better person inside out,

You made me want to prove that immigrant mothers can achieve and contribute to society,

You made me want to prove that nothing is impossible if you're dedicated to,

You made me want to prove that working hard pays off!

You made me want to prove no one can stop you from doing what you want!

Thank you, Jiyul, for co-designing StoryAI, playtest it whenever I ask you to, and critique harshly on what you like and what you don't like!

Thank you, Sunyul, for understanding that mommy always stays in front of the computer and busy but you're the sweetest girl, my healing, my multi-vitamin!

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

My deepest appreciation goes out to my committee chair, Professor Kylie Peppler, for her patience, care, and support throughout my doctoral journey. Her guidance and help gave me the strength to pursue interdisciplinary research while maintaining my identity as a designer. I would like to thank my committee members, Professor Katie Salen Tekinbaş and Professor Kurt Squire, for their valuable feedback and guidance over the past few years as I was developing and refining this new research topic.

This work would not be possible without my academic brother, amazing programmer, Shenshen Han's dedicated contribution to translating the unstoppable wild ideas into an actual, functional prototype. Kylie Peppler and Seth Corrigan's guidance on StoryAI helps us to achieve beyond what we can achieve. StoryAI's excellent teamwork made it possible to win the VITAL Challenge as a semi-finalist.

Many thanks go to my friends, colleagues, and collaborators at UC Irvine and beyond. My position as an informatics student became a great hybrid into educational informatics by working at the Creativity Labs with amazing colleagues and mentors from the School of Education. I appreciate Joey Huang's support since the beginning of my PhD journey helped me develop multi-disciplinary research experience. Nikki, Santiago, Lora, Phebe, and Aakriti, my lab friends, gave me the confidence to improve my work and fill out my weaknesses. I appreciate Zhenyao for reaching out and helping me start the topic of AI and education. The endless discussion we had inspired me greatly. Additionally, I would like to thank Kunlie He for having lunch with me at UTC and sharing such insightful research experiences and ideas. I look forward to collaborating with these folks in the future! Also my Korean friends at Informatics, eating K-food with you all were just stress relievers for me!

I am also extremely grateful for my funders, the UCI Beall Applied Innovation Center, and the VITAL Challenge Award powered by NSF and Digital Promise. I also want to thank Daniel Epstein for his outstanding guidance on academic publication, and how to critique and construct academic papers. As well as Aure Schrock's support of the papers, the feedback and suggestions they provided were critical to the publication process.

Last but not least, I want to thank my family who always accompany me and cheer for me. My husband is the most dedicated and caring person–he trusts me, values my passions, and works side-by-side with me to balance family and school responsibilities. My parents and in-laws for their unconditional love and support, which was all of my foundation for me to pursue what I pursue right now. I am grateful for their selfless support that has helped me complete this journey and given me the courage to embark on my academic journey.

# VITA

## Ariel Han

### EDUCATION

**Ph.D. in Informatics** **2024**
University of California, Irvine *Irvine, CA*

**M.S. in Entertainment Technology** **2013**
Carnegie Mellon University *Pittsburgh, PA*

**B.A. Information Technology, Industrial Design, Fine arts** **2011**
Seoul National University *Seoul, South Korea*

### HONORS AND AWARDS

**VITAL Prize Challenge Award** **2023**
University of California, Irvine *Irvine, California*

**UCI Beall Applied Innovation** **2023**
University of California, Irvine *Irvine, California*

**Doctoral Consortium (Interaction Design and Children)** **2023**
University of California, Irvine *Irvine, California*

## PUBLICATIONS

### Referred Conference Proceedings

**Han, A.,** Zhou, X., Cai, Z., Han, S., Ko, R., Corrigan, S., & Peppler, K. 2024. Teachers, Parents, and Students' Perspectives on Integrating Generative AI into Elementary Literacy Education. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '24, May 11–16, 2024, Honolulu, HI, USA: Association for Computing Machinery. https://doi.org/10.1145/3613904.3642438

**Han, A.,** Corrigan, S., Han, S., & Peppler, K. (2024). Co-Design a Logic Model for Inclusive AI-Powered Learning Application with Primary School Teachers. International Society of the Learning Sciences. ISLS Proceedings: 17th Annual Conference of the Learning Sciences (pp. X-X). Buffalo, New York: International Society of the Learning Sciences.

**Han, A.,** & Cai, Z. (2023). Design implications of generative AI systems for visual storytelling for young learners. Interaction Design and Children https://doi.org/10.1145/3585088.3593867.

**Han, A** (2023). Implications of AI art generators to broaden visual literacy and creative expression for young learners International Society of the Learning Sciences (ISLS) Annual Meeting 2023. International Society of the Learning Sciences.

**Han, A.,** Keune, A., Huang, J., & Peppler, K., (2022). Visualizing Family Engagement in Museum Settings. In: J. Oshima, T. Mochizuki, & Y. Hayashi (Eds.) International Collaboration toward Educational Innovation for All: International Society of the Learning Sciences (ISLS) Annual Meeting 2022 (pp. 1094-1095). Hiroshima, Japan: International Society of the Learning Sciences.

Huang, J., **Han, A.,** Sedas, M., Telfer-Radzat, K.,& Peppler, K., (2022). Crafting paper circuits: Gendered materials for circuitry learning. In J. Oshima, T. Mochizuki, & Y. Hayashi (Eds.) International Collaboration toward Educational Innovation for All: International Society of the Learning Sciences (ISLS) Annual Meeting 2022. Hiroshima, Japan: International Society of the Learning Sciences.

### Conference Workshop Papers/Doctoral Consortium

**Han, A.,** Cai, Z., Jeong, S., & Choi, S. M. (2023). AIStory: design implication of using generative arts AI for visual storytelling. Child-Centered AI Design: Definition, Operation, and Considerations ACM CHI 2023 Workshop.

**Han, A.** (2023). Ai Virtuous Circle: Preparing Youth for the future of Creative Economy. Proceedings of the 22nd Annual ACM Interaction Design and Children Conference. https://doi.org/10.1145/3585088.3593919

**Han, A.,** Huang, J., Villanueva, A. M., Peppler, K. A., Liu, Z., Zhu, Z., & Ramani, K. (2022). Coding a MacGuffin: Recommendations for Teaching Narrative-based IoT De-

sign. In Proceedings of the 2022 American Educational Research Association (AERA). https://doi.org/10.3102/2009108

**Journal Articles**

Lee, U., **Han, A.,** Lee, J., Kim, J., Lee, E., Kim, H., & Lim, C. (2023). Prompt Aloud!: Incorporating image-generative AI into STEAM class with learning analytics using prompt data. Education and Information Technologies. https://doi.org/10.1007/s10639-023-12150-4

Huang, J., **Han, A.,** Villanueva, A. M., Liu, Z., Zhu, Z., & Ramani, K. Peppler, K., A., (2023). Deepening Children's STEM Learning through Making and Creative Writing. In Proceedings of the 2023 International Journal of Computer Child Interaction, IJCCI. https://doi.org/10.1016/j.ijcci.2024.100651

Peppler, K., Keune, A., & **Han, A.** (2021). Cultivating data visualization literacy in museums. Information and Learning Sciences, 122(1/2), 1–16. https://doi.org/10.1108/ILS-04-2020-0132

# ABSTRACT OF THE DISSERTATION

StoryAI: designing, developing, and evaluating Generative AI-powered story-authoring platform for young learners

By

Ariel Han

Doctor of Philosophy in Informatics

University of California, Irvine, 2024

Professor Kylie Peppler, Chair

The dissertation consists of three studies that are the process of designing, developing, and evaluating generative-AI-powered story-authoring platforms for children. The first study focuses on the formative study on how stakeholders in education (i.e., teachers, parents, and students) perceive and leverage generative AI platforms (i.e., ChatGPT and Stable Diffusion) for writing activities. I found that the GenAI 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. I provided 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.

The second study focuses on co-designing a logic model in informing designing AI-based Writing Tutoring Platforms (AWTP) with educators, then designing and evaluating AWTP prototypes that focus on opinion writing. From the co-design process, we identified the platform's potential users, features, functionalities, and desired outcomes. With this insight, we created a prototype, AWTP. The usability study findings with the AWTP prototype suggested AWTP's efficacy in improving students' writing engagement by increasing their time

spent in writing, total word count, and lexical diversity. Feedback study revealed AWTP's potential efficacy in improving motivation in writing by reducing anxiety over writing for emergent writers.

The third study focuses on designing, developing, and evaluating story-authoring platforms, StoryAI, for narrative writing for children. From the usability study, I found StoryAI's efficacy in students' perception of writing competencies (i.e., planning, translating, and revising) as well as AI literacy (i.e., perception, confidence, and motivation).

Overall, the three studies provide convincing evidence for leveraging GenAI-based learning platforms to support children's literacy development. The findings are intended to contribute to designing interactive intelligent systems for learners and educators, leveraging story creation to teach various subjects including writing on different topics (i.e., science, history) to empower children's long-term academic success and development.

# Chapter 1

# Introduction

Storytelling is an effective pedagogical strategy that can be incorporated into lessons to increase students' competencies across various disciplines [137, 142, 123, 84]. Due to the interrelated nature of the processes involved in reading and writing and social interactive elements in storytelling among listeners and tellers, storytelling has advanced participation and engagement in learning experiences [51, 201, 191]. Digital Storytelling platforms offer various advantages to promote learning not only literacy (i.e., reading and writing), but also abilities to convey meaning to others, convey information, and express their thoughts via interactive format with various media such as image, video, and text [95]. However, due to the individualized nature of storytelling activity involving a listener and a teller pair, it is hard to facilitate one-on-one interactions for all learners in educational settings. To benefit more students from storytelling activities, text-based, and voice-based conversational agents have been utilized [191, 93, 108, 127] that can engage in meaningful conversations, as it does not require the presence of another person.

The text-based conversational agent, a chatbot, has been studied and utilized to improve personalized and adaptive learning across a broad range of educational contexts, including lit-

eracy [57, 201, 189], mathematics [185, 199], and system thinking [127]; it has demonstrated effectiveness not only in learning itself but in advancing motivation by reducing anxiety [157, 110]. However, most previous works involved rule-based chatbot systems, which were limited in their scope of interaction [158, 94]. The rapid progress in generative AI (GenAI) technologies, exemplified by innovations like ChatGPT, and Stable Diffusion, has opened up unprecedented opportunities for dynamic conversations and creative activities, capturing significant interest within the field of HCI and educational research [106, 82]. Furthermore, the advent of Large Language Model (LLM) technology has significantly enhanced the capabilities of these AI agents, enabling them to deliver responses that are remarkably human-like and replete with accurate information [82]. Historically, chatbot applications have utilized straightforward, textual interfaces to afford users the ability to retrieve information, engage with services, or partake in entertainment via online messaging platforms [158]. More recent efforts in the field have aimed to expand the utility of chatbots, employing them to assist individuals facing social and communicative challenges, or to encourage users to pursue domain-specific learning objectives, such as language acquisition and storytelling [57]. The preference for chatbots as a medium of informational interaction stems from their ease of use, naturalness, and intuitive design.

This dissertation consists of three papers focusing on the processes of designing, developing, and evaluating GenAI-based story-authoring platforms that are designed to serve as a learning companion during story-writing activities. Each of the papers addresses the progress started from the formative study (i.e., need findings), co-design features and functions, and iteratively design and develop, to evaluate the effectiveness of the platform. They provide a comprehensive investigation of what educators and students need, perceive, and utilize GenAI-powered, text-based conversational agents.

## 1.1 Dissertation Overview

This dissertation consists of five chapters: an introductory chapter, three individual study chapters, and a concluding chapter. Instead of having separate chapters for literature review and methodology, each study chapter is self-contained, containing its own literature review and method section tailored to the specific context of the study.

Study 1 focused on understanding stakeholders' perception of generative AI technology and its potential usage in writing education.

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

1. 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?
2. What are the values and motivations towards GenAI that differ among stakeholders in education?
3. In what ways can GenAI systems be designed so that they are effective, engaging, and safe for teaching literacy to 2nd to 6th graders?

Results revealed that stakeholders perceive GenAI systems differently: 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, I highlight possible obstacles and concerns regarding authorship and ownership issues with writing outputs, challenges of examining students' agency in learning, and difficulties in controlling bias and hallucinated content created by GenAI systems. Based on the findings, I provide design implications to mitigate the weaknesses 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 GenAI 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 GenAI-LLM systems that allow teachers to curate child-AI interaction.

Study 2 intended to co-design a logic model, a conceptual framework that specifically highlights the causal pathways that guide design decisions on designing key features and functions to achieve the desired outcomes. To progress from overarching design principles to more specific pedagogical strategies for our GenAI-powered educational tool, I collaborated with classroom teachers to develop logic models – and design a high-fidelity prototype for an opinion writing module called AI-based Writing Tutoring Platform (AWTP) This process was not just about the technical design of AI tools, but also about aligning these tools with the educational goals and practices of teachers. Two questions were asked in this study:

1. To what extent can AI-based writing platforms help educators achieve their goals and objectives?

2. How does the AWTP support the efficacy of writing for children?

In summary, co-design study uncovered the blueprint of AWTP for writing support for children: ***Context/Users*** (struggling writers and low-level readers), ***Inputs/Activities*** (playful literacy activities, text chats, culturally responsive imagery selection, writing models, feedback, translanguaging, and writing strategies), ***Outputs*** (increased amount of time spent in writing, word count, genre-specific writing strategies), and ***Outcomes*** as fluent in spelling, typing, word processing, lexical diversity, syntax, and verbal reasoning, ultimately increase engagement and motivation in writing. These can be facilitated by developing AI-powered story-authoring platforms through conversational design, chatbot systems, and customizable editing stations.

The usability study with AWTP revealed that t students stayed in writing longer when they utilized AWTP for their writing, the average writing time for students without AWTP was 26.15 minutes and they stayed 10.62 minutes longer when they wrote with AWTP. Additionally, students tended to write more (total word counts) with AWTP as they wrote an average of 162.62 words but they wrote 210.77 words when they wrote with AWTP which indicated the AWTP's potential to promote writing engagement. We also noticed students' lexical diversity in their writing has increased.

Study 3 focused on the design, development, and evaluation of StoryAI, a story-authoring platform for children specific to narrative writing. This study examined children's writing efficacy in planning, translating, and revising as well as AI literacy specific to prescription, confidence, and motivation. Additionally, I examined the usability of StoryAI in understanding ownership over the final writing output, enjoyment, ease of use, sense of collaboration, and satisfaction. Four questions were asked in this study:

1. In what ways does StoryAI help students enhance their writing efficacy and motivation?
2. Would writing with StoryAI help students develop AI literacy?
3. In what ways do students make use of StoryAI to support their writing?
4. How do students perceive StoryAI AI agents for their writing?

The results from the usability study on the writing workshop with StoryAI indicate notable improvements in students' perceptions of their writing efficacy across three key competencies—planning, translating, and revising—from the pretest to the post-tests. Furthermore, the data revealed that students' AI literacy has advanced across three dimensions: awareness of what AI is, confidence in using AI, and motivation to integrate AI into their daily lives [126].

In these lines of research, empirical evidence will be provided on the effectiveness of GenAI in educational settings, allowing a comprehensive understanding of its advantages and limitations. This evidence can guide future GenAI integration in educational contexts. The research will enhance our understanding of child-centered AI by underscoring the need to develop technologies tailored to the cognitive and emotional needs of young learners. This emphasis not only sheds light on new research avenues within GenAI and education but also paves the way for future innovations that are specifically designed for children.

In the concluding chapter, I explore the broader implications of the findings derived from the three studies and propose directions for future research.

### 1.1.1 Theoratical Background

The project, StoryAI, is grounded in the following theories: constructivism, the cognitive theory of writing, and Vygotsky's social constructivism [139, 173, 130, 172, 62]. Constructivism highlights active learning where learners construct meaning through active engagement in their learning processes [139]. Vygotsky (1978) highlights that learners construct knowledge through social interaction. His concept of the Zone of Proximal Development (ZPD) underscores the critical role of guidance from more knowledgeable individuals in the learning process [173]. The cognitive theory of writing [62] identifies three key dimensions: the task environment, which includes external factors like audience, purpose, and assignments; the writer's long-term memory, encompassing prior knowledge and experiences that influence writing; and the recursive nature of writing processes, involving planning (idea generation), translating (text composition), and reviewing (revising) [62].

The integrated theoretical background, emphasizes how the cognitive theory of writing, constructivism, and Vygotsky's ZPD and social constructivism collectively provide a comprehensive framework for understanding and enhancing the writing process with the GenAI-based

story authoring platform. These theories collectively inform the design and functionality of a GenAI-based story authoring platform that can provide personalized feedback and guidance, facilitate social interaction and collaboration, and support recursive writing processes, thereby enhancing the user's writing skills.

First, constructivism's active participation in learning applies to Flower and Hayes' cognitive processes where the writer builds and refines their understanding of the topic through active practices of planning, translating, and reviewing. This aligns with the constructivist view that knowledge is actively constructed by the learner. Second, prior knowledge and experiences, in writing, ones' prior knowledge influences how writers plan, generate ideas, and make revisions. Constructivism also emphasizes that new learning is built upon the foundation of existing knowledge, making the writer's background essential in the writing process. Third, problem-solving and critical thinking, as writing is a complex problem-solving activity that requires planning and decision-making. Constructivism supports this exploration, inquiry, and reflection, which are essential in the writing process. Fourth, is the recursive process of learning and writing, as these theories recognize that learning and writing are not linear activities, rather, they involve ongoing reflection, revision, and refinement. Lastly, in a social context and collaboration, writing is influenced by the social context, including the intended audience and the feedback received from others. Constructivism supports collaborative learning environments where social interaction and discussion help individuals construct knowledge which includes, scaffolding in writing, the importance of guidance and support from knowledgeable individuals, and co-construct knowledge through social interaction and shared experiences [130].

By weaving together these theories, I aim to provide a holistic approach to designing AI agents that support writing instruction, emphasizing cognitive processes of writing, active learning, social interaction, and scaffolding. Instructors can create AI-driven conversations that guide and provide feedback to students, facilitating social interactions where students

collaboratively write with AI agents. Additionally, this method promotes the active construction of knowledge, as students engage dynamically with writing processes through conversations with AI agents.

# Chapter 2

# Perspectives on Generative AI into Literacy Education

## 2.1 Study Abstract

The viral launch of new generative AI (GenAI) 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 GenAI 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 GenAI systems for learning and teaching writing. We found that the GenAI 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.

## 2.2 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 [148]. 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 [101, 149]. 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 [44, 144]. The ongoing discourse in education around generative AI (GenAI) emphasizes the need for comprehensive research into its integration within educational contexts [160].

GenAI, also known as Generative Adversarial Networks (GANs), GenAI systems have GenAIned significant attention within the HCI community [124, 175, 196]. 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 [12, 33, 131, 122], including generating novel outputs-such as images, text, music, or video-based on patterns it learned from large datasets during its training [26]. The HCI research community has started to examine utilities and interaction techniques with these systems [104, 184], focusing on new interaction styles [197, 87], Large Language Models' (LLM) capacities [104], and how to adapt the systems to creative activities for adults [112, 69]. While the advancements in GenAI 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 [144]. Recognizing the role of storytelling in child development [88] and its impact on critical skills like imagination and comprehension [115], it becomes clear that integrating such advanced technologies in education demands a careful balance. This approach should respect both the potential of GenAI 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 GenAI 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 GenAI 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 GenAI-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 GenAI systems differ among stakeholders in education?

- In what ways can GenAI systems be designed so that they are effective, engaging, and safe for teaching literacy to 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 GenAI in writing projects. In total, we report on insights from 40 participants who present unique perspectives on GenAI from three groups of stakeholders in education (i.e., 16 teachers, 12 parents, and 12 students).

From the study, stakeholders' perceptions towards GenAI 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 GenAI 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 GenAI 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 GenAI-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 GenAI in education and by offering insights that can guide the design and implementation of GenAI 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 GenAI 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 GenAI 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.

## 2.3 Literature Review

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.

### 2.3.1 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 [144]. To contextualize our investigation of the potential applications and benefits of emerging GenAI 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 [144]. Therefore, Reich 2020 argues that educational innovations must be deeply rooted in the realities of teaching and learning.

Reich [144]'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 (GenAI) 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 [109]. As GenAI 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. [160], are beginning to explore GenAI'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.

### 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 [192, 186, 21, 138, 90, 163, 34, 128], yet little of this prior work addresses directly how AIED integrates holistically into educational settings [34, 89]. An exception to this is Chiu et al. [37] 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. [37] 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 [170] among educational stakeholders, including teachers [107, 140], children [187, 14, 194, 28, 182], and parents [169, 65, 190, 64]. 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. [107] 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 [200]. 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 [182]. Additionally, Xu and Warschauer. [194] 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 [194].

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.

**Emerging Trends and Challenges in Generative AI Applications for Education**

The rapid advancement of GenAI, such as large language models (LLMs) and Text-to-Image (TTI), learn patterns and structures from existing data and generate new content [179]. 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 [132]. Ahmad et al. [10] examined the implications of ChatGPT in the education sector, emphasizing the need to develop skills for using LLMs and GenAI 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 [10].

Research on AI systems in literacy education (reading and writing) focuses on LLM-based chatbots for language learning [201, 9], scientific writing [67], creative writing [198, 39, 156], and creative storytelling [81, 200]. For example, Gero et al. [67] studied how LLM-powered co-writing platforms can enhance engagement and idea generation with STEM graduate students. Yuan et al. [198] 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. [104] 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. [198] and Lee et al. [104] also raised questions about writer's feeling of ownership over their writing outputs and indicated the results were uncertain.

Recent GenAI-powered educational applications offer potential opportunities to leverage GenAI systems in teaching and learning (GPT-3, TTL) [30]. For example, Speak [13] uses GenAI 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 [117, 25, 155]. MagicSchool.ai [11] is a web application that uses GenAI 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 [102].

However, it is still unclear how these new interactions, user experiences, and learning engagements affect learning outcomes [18]. The current story creation apps powered by GenAI systems produce whole stories for students, which raises a question about whether it could promote language learning or undermine creativity [81]. Hence, further research is needed to ensure that GenAI-powered learning tools are effective and age-appropriate.

### 2.3.2 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 [66]. 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 [188]. Enhancements are also needed for human-AI collaboration to relieve the repair burden on families during their communication breakdowns with CA [21]. As for children's perspective, research efforts have been made to investigate children's perception of their data utilized online [176], children's autonomy over the technology [177], and AI technologies' influence on child development. With the recent advancement of GenAI, 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 [152], and in turn,

challenges AI literacy education by equipping children with some basic AI literacy in both classroom and family scenarios [161, 56]. 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 [198], creative writing [59], and drawing [200]. Wordcraft [198] 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 [200]. 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 [59]. 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 [129, 162], where creativity requires writing with a relevant purpose, understanding, judgment, and evaluative abilities in ways that are deemed original and valuable to a community [43]. 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 [63]. 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.

## 2.4 Method

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 GenAI systems may provide design implications to help guide the development of GenAI tools and systems for elementary school students. To elicit stakeholders' perspectives on potential possibilities and limitations of a GenAI-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 [52]) and a chatbot powered by LLM (i.e., ChatGPT [32]). 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 GenAI platforms.

### 2.4.1 Study Procedure

#### Workshop with Families

In April 2023, we conducted a workshop with families with children ages 8 to 12 (2nd and 6th graders) (Table 3.1) 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

Table 2.1: 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 GenAI tools (ChatGPT and Stable Diffusion) |
| 30 min | Let's use generative AI to write a visual story |
| 15 min | Share your story (Reflection) |

critical importance of this phase in developing reading and writing skills [78]. 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[75]. Given this, our study aims to investigate how enhanced engagement with writing and literacy activities facilitated by GenAI platforms can positively or negatively impact learning in these formative years.

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 [174]. It was the first time the students had used GenAI-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 [145]. Given the California State Standards in elementary literacy education, we chose narrative writing activities for students [73] 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 [145]: 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

Table 2.2: 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 |

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.

**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 3.2).

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 GenAI systems (i.e., features and functionalities) and asked about their 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 GenAI systems (Midjourney), whereas the rest were unfamiliar with these systems. Teachers who have used GenAI 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 GenAI 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 GenAI 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 GenAI systems in your class or for yourself?"),

- Their opinions of their intended usage of the GenAI systems, and their opinions and concerns about them (e.g., "Can you tell me your thoughts about the GenAI systems as students use them for writing?", "Can you share your opinions on whether or not

Table 2.3: 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 |

GenAI 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).

### 2.4.2 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 [166, 151]. We

conducted an inductive approach to analyze interview data [165]. 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 [116]. 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 GenAI; 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.

### 2.4.3 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 GenAI 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 system-wide 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.

## 2.5 Result

By analyzing qualitative interviews and observational notes, we uncovered multiple perspectives regarding the use of GenAI in literacy education. In this section, we report major

findings regarding our participants' opinions and experiences with GenAI. We outline the values and perceptions of multiple stakeholders (see Figure 1), then elaborate on the findings in the advantages and constraints of GenAI 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 GenAI in literacy education, particularly teaching and learning writing. GenAI 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.

### 2.5.1 Multifaceted Views on the Role of GenAI in Literacy Education

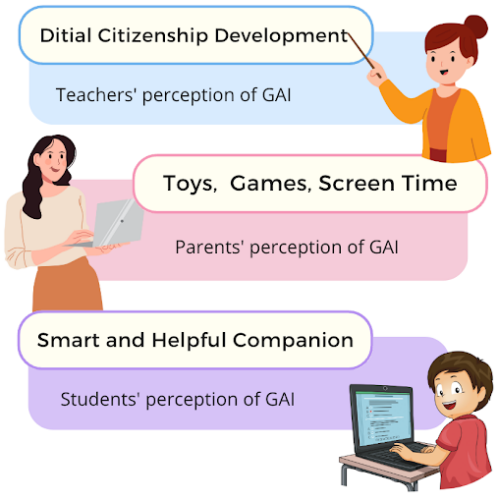


Figure 2.1: Summary of each stakeholder's perspectives and opinions of GenAI systems (top: teachers, middle: parents, bottom: students).

**Adapting Digital Transformation: Teachers' Perspective on Integrating GenAI 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 GenAI 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 GenAI 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 GenAI systems as another tool that they will need to learn how to use.

Teachers pointed out that GenAI systems can also be used to support educational processes [38], nine respondents (9/16) emphasized that GenAI 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 GenAI–LLM chatbot lacks the capacity to be fully integrated into human conversations but can be useful for brainstorming ideas:

*"I don't think Al 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 (GenAI) into their teaching pipeline along with digital literacy development. In addition, they stressed the importance of equipping their students with the ability to use GenAI systems to develop their digital citizenship.

**Parental Caution: Attitudes Toward GenAI 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 GenAI 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 GenAIs 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 GenAI 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 GenAI 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 [141]. Most parents (8/12) perceived the GenAI systems as new to them, so they had difficulty imagining how it would affect young minds. For example, P03 and P04 mentioned:

*"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

GenAI systems properly, most parents prioritized promoting critical thinking and problem-solving instead of introducing GenAI systems to their children. Moreover, parents (n=11, mothers) presented anxiety over adapting GenAI 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 GenAI 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).

### Creative Allies with Caveats: Students' Mixed Perceptions of GenAI 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 GenAI in the process of creative writing, with (10/12) of students pointing out the efficiency of using the GenAI-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 GenAI had not generated the content they intended. As a result, we concluded that instructing and teaching prompt writing would enhance efficiency and adaptability [111]. Second, we found that the randomness of the output generated by GenAI 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.


Figure 2.2: Summary of our findings of potential affordances and limitations of GenAI systems for writing projects in elementary school settings

### 2.5.2 Delineating Advantages: GenAI's Contributions to Literacy Education

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

**Enhancing Pedagogical Efficiency: GenAI in Crafting Customized and Scaffolded Mentor Texts**

The teachers (16/16) all affirmed that GenAI 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 GenAI 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 GenAI 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 GenAI 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 GenAI-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 GenAI systems to generate scaffolded mentor texts and teaching materials for different levels of students' capacity.

### Scaling Individual Attention: GenAI 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 GenAI 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 GenAI 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 GenAI 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 GenAI systems conversation.

**Culturally Inclusive Pedagogy: GenAI's Capabilities for Culturally Relevant Literacy Feedback**

The other aspect of using GenAI for personalized learning is to provide culturally relevant feedback and ideas [133]. Teachers and parents were particularly intrigued about the possibility of translating languages and providing examples of different cultures with GenAI systems. Teachers intend to utilize GenAI 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 aGenAIn, 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 GenAI 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 GenAI 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.

### 2.5.3 Navigating the Gray Areas: Challenges and Constraints of GenAI in Literacy Education

Our research indicates that GenAI 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 GenAI systems to generate biased or inaccurate content stemming from their inherent randomness and uncertainty.

#### Ethical Quandaries, and Accountability in GenAI-LLM Writing systems

Nine teachers (9/16) expressed concerns about introducing GenAI 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 GenAI 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 GenAI 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 GenAI 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.

### The Agency Dilemma: Unpacking Student Agency in the Complex Role of GenAI in Student Literacy

There has been difficulty determining the level of agency students have over their writing outputs when using GenAI, particularly when it comes to disambiguating how much students write (i.e., the ideas, the sentence, the paragraph, the word choice) versus what GenAI suggests and generates. From the writing workshop, we observed that many students (8/12)

just copied and pasted directly from GenAI-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 GenAI 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.

**Erratic Outputs: Limitations and Concerns in Deploying GenAI for Literacy Education**

Like any generative AI chatbot or voice assistant–such as Siri, Alexa, or Microsoft's ill-fated Tay [171]–some individuals intentionally try to corrupt or manipulate GenAI-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 [134, 150, 60]. Less malicious but still disruptive are instances where a GenAI system produces surreal or nonsensical responses to user prompts. GenAI 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 [24, 31]. 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 GenAI 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 GenAI-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 GenAI 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 GenAI 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.

## 2.6 Discussion

From the study, we examined the potential advantages and challenges of using GenAI 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 GenAI 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 GenAI learning systems. Additionally, we outline three key design considerations essential for developing effective GenAI-based educational applications.

### 2.6.1 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 [44]. 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 [44]. 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 [144].

The recent LLMs have brought breakthroughs of open-ended conversational systems, which perform open-domain dialog with any topics [86] and it offers the capability to be fine-tuned [135], enhancing its performance to align with specific domains and instructional objectives

[178, 202]. Unlike traditional MOOC platforms, which rely on human-guided instructions, educators can now train the LLM with specialized datasets and employ prompt engineering techniques [180] 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 GenAI systems can be integrated with new pedagogical approaches.

Consequently, new breakthrough systems like GPTs [132] 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 [91, 15]. 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.

**Double-Edged Sword of GenAI in Education**

From the study, we found educators were drawn to use the GenAI 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 GenAI systems' capabilities to facilitate interest-driven learning, particularly about

culturally relevant approaches in writing projects [136, 17].

The use of GenAI 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. GenAI 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 GenAI systems can lead to potential challenges, including the development of biased perspectives and the generation of inaccurate or 'hallucinated' results [55]. Such issues underscore the critical need for careful moderation and strategic oversight, such as the implementation of customized models [50] (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 GenAI while minimizing its risks for educational settings.

### 2.6.2 Recommendations for System Designers and Developers

As part of this discussion, we propose the design considerations of GenAI-powered writing platforms to inform the designing of safe and accessible GenAI systems for elementary school settings. To capitalize on the perceived benefits of educational uses of GenAI while mitigating the concerns from our stakeholder groups, educational GenAI platforms should: 1) provide guardrails to protect students' authorship issues in GenAI-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 GenAI systems.

### 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 GenAI generates the majority of the content for students [105, 40, 45]. Even though studies have examined GenAI-LLM-powered writing systems, such as Gero et al. [67], Lee et al. [104], and Yuan et al. [198], 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) [168]. Gero et al. [67], and Lee et al. [104] 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 [49, 99, 100].

To better understand what guardrails and guidelines need to be implemented into the development of GenAI-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 GenAI-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 [92]. This method allows educators to compare student writing with AI output, offering insights into the extent of AI reliance on student work.

**Enhancing Student Agency through Role Allocation in GenAI Systems Design**

We observed that when students encountered open-ended GenAI 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 GenAI-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 [136, 58].

According to the workshop with students and teacher interviews, promoting students' agency as writers is essential [103], 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 GenAI 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 [195, 14]. 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. [128] discusses the benefits of designing prompts that enable chatbots to foster systemic thinking (such as idea generation and questioning). Specifically, Nguyen. [128] 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 ex-

pert 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.

### Balancing Flexibility and Control GenAI-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 GenAI LLM systems, it is essential to design a system that balances flexibility and control with adults-in-the-loop systems [118, 80, 201]. Yuan et al. [198] 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 [115].

Hence, we propose designing an 'educators' view' that allows educators and/or parents to easily 'prompt' and curate GenAI-based chatbots' conversation to facilitate a secure mode of student-AI interaction for writing. For example, the new systems will allow educators to prompt GenAI systems to carry on their lessons, similar to the current tool that designs a chatbot with flow-based interfaces, such as Voiceflow [16]. Our suggestion is to develop flow-based interfaces [61] (or block-based interfaces [23]) 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 GenAI-LLM-based chatbot might have and provide open-ended flexibility, with low floors and high ceilings [147].

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 [154] to understand their unique languages and mental model to create an appropriate conceptual model that aligns with educators' goals [107]. 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.

### 2.6.3 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 GenAI utilization. This should involve a comprehensive analysis of system logs and behavioral data within GenAI platforms. This includes leveraging GenAI 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 GenAI tools influence user interactions, experiences, and learning outcomes over time.

To promote accelerated learning through GenAI-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 [72, 121, 71]. 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.

## 2.7 Conclusion

In this paper, we explored the stakeholders in education's perceptions and opinions regarding the advantages and limitations of leveraging GenAI 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 GenAI 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 GenAI 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, GenAI, and NLP communities to design a GenAI-powered platform for literacy education.

# Chapter 3

# Co-Designing an AI-based Writing Tutoring Platform (AWTP) with Primary School Teachers

## 3.1 Study Abstract

This study explores the design and development of an AI-based Writing Tutoring platform (AWTP) by involving teachers to co-design a logic model to guide the design and implementation of important features and functions tailored for young learners (ages 8 to 12). Amidst the rise of Generative AI tools, there's a need to bridge the gap between AI advancements and their practical use in K-12 writing education. To address this, the researchers collaborated with primary school teachers and synthesized insights from the literacy education literature. From the co-design process, we identified the platform's potential users, features, functionalities, and desired outcomes. With this insight, we created a prototype, AWTP. We conducted a user study with thirteen students as well as a feedback study with fifteen

educators to understand the feasibility, applicability, and efficacy of AWTP in writing education. The findings suggested AWTP's efficacy in improving students' writing engagement by increasing their time spent in writing, total word count, and lexical diversity. Feedback study revealed AWTP's potential efficacy in improving motivation in writing by reducing anxiety over writing for emergent writers. Lastly, We offered design insights for AI-based learning tool designers and developers focusing on both students and teachers.

## 3.2 Introduction

The ability to write is critical to success in a wide range of circumstances and career paths. Writing provides the fundamental abilities to articulate ideas, argue opinions, and synthesize multiple perspectives [77]. As a result, proficient writing is essential to effectively communicating with others, including teachers, peers, colleagues, coworkers, and the community [42, 98]. Thus, proficient writing is essential not only for academic achievement but also for professionals to manage their daily duties and to foster career advancement. [78].

Despite clear evidence highlighting the importance of writing, teaching writing remains a complex task, with many students struggling to develop strong writing skills. [74]. Because writing competence is not acquired quickly, it takes many years to develop [74] and requires guidance and practice [79]. It is even more difficult for newly emerging writers to develop their writing skills because they have not yet mastered fundamental writing skills for expressing their ideas and translating them into sentences. In elementary school, students are still learning strategies and schemas for planning, drafting, revising, and editing their texts. Students are still developing topic knowledge for future writing projects as well as learning about text characteristics and good writing. More importantly, research has shown that there is a lack of teachers' education resources in teaching writing [70], and teachers have a hard time providing appropriate and personalized writing education in school settings [78].

Developing AI-assisted writing tutoring platforms can serve as a supportive tool for teachers, enabling them to integrate these technologies into their lessons and practices. However, a significant gap persists in understanding teachers' actual needs and objectives in their unique settings among educational technology designers and developers. The recent launch of Generative AI (i.e., ChatGPT, Dall.E3) sparked educational discourse, raising both fears and hopes of adapting the cutting-edge platforms in education [82, 132, 96]. However, there is a gap in the field's understanding of how to best leverage AI technologies for improved students' writing efficacy. As writing processes are complex, and education is deeply rooted in the quality of pedagogies, not just the technical innovations.

We utilize a combined backward design and co-design approach in concert with logic model development to bridge this gap. We co-design a logic model as a collaborative effort of six educators (K-12 classroom teachers) to identify who might be users what context these new systems can fit in, and what features and functionalities are needed for desired outputs for an AI-based writing tutoring platform. We recorded co-design sessions and analyzed the resulting qualitative data to elicit teachers' intended goals, needs, and suggestions. We also drew insights from synthesizing several lines of scholarly literature including literacy education to identify desirable outputs (what learners achieve) [46, 78, 73].

A key part of our strategy involves collaborations with primary school teachers to build logic models to specify the desired input, activities, outputs, and outcomes associated with the technology. Through the process, teachers' expertise in existing strategies for teaching and learning literacy skills is integrated with the development team's understanding of the affordances and limitations of current AI platforms. The design process incorporated classroom teachers to understand their unique perspectives and needs. Furthermore, we incorporated current research on literacy education to inform the final inputs and outputs of the resulting logic model.

With this insight, we created a prototype, an AI-based writing tutoring platform (AWTP)

which includes text-based chatbot systems in which AI agents guide students through the writing process including narrative writing and opinion writing. We conducted a usability study to examine the efficacy of the AWTP in writing education with thirteen students ages 8 to 10 who focused on the opinion writing module. We analyzed students' writing output and measured writing time in minutes, total word counts, and lexical diversity. The data revealed that students' writing more (total word count) has increased as well as their time spent in writing (see Figure 3.2). Also, students' essays have higher lexical diversity when they write with AWTP. Following the writing workshop, we conducted a feedback study with fifteen educators to understand the applicability and feasibility of the AWTP from educators' perspectives, they provided useful feedback where we categorized the themes into affordance, concerns, and suggestions which we explained in more detail in the finding section (Section 3.8.3).

The contribution of the study to the learning science community is two-fold: first, we collaborated with classroom teachers (K-12) to understand their opinions, and to develop a logic model to inform designing learning applications tailored to their specific needs and motivations. As a first step towards designing and developing a teacher-centered AI-powered educational platform that meets their unique needs. This collaboration underscores an opportunity to foster joint efforts between researchers, educators, AI, and NLP specialists to design an AI-powered platform for K-12 education. Second, we present the process of constructing a causal pathway, incorporating strategies to achieve desired outcomes, in the creation of innovative educational technology, focusing on the development of creative AI-powered learning tools—a logic model that offers valuable insights to learning designers and researchers alike to find features and functionalities of integrating such emerging tools for educational purpose.

## 3.3 Literature Review

### 3.3.1 Logic Model

A logic model is a conceptual framework that specifically highlights the causal pathways leading to desired outcomes [159]. The logic model guides the big picture of how key features are intended to achieve the goals of a given product or intervention. Logic models have been a long-term feature in educational research and design [159]. Their utility derives from their ability to systematically map out the expected functionality and impact of a product and guide design decisions. Logic models identify the expected context and users, inputs, and activities (i.e., features and functionality), and describe how these influence the user's experience, the users' outputs (i.e., user experience data), and the expected outcomes (i.e., potential benefits of using the product). Logic models can provide valuable guidance for developing innovative educational technologies by clarifying a technology's target audience, functionality, and desired impacts used in the context of a backward design framework [181]. The purpose of the current study is to report on the development of a logic model that informs the design and development of a cutting-edge AI-powered story-authoring platform for young learners.

### 3.3.2 AI-powered Literacy Development

The rapid advancement of Generative Artificial Intelligence (GenAI) models, such as Large Language Models (LLMs) and Text-to-Image (TTI), make it possible for them to learn patterns and structures from existing data and to generate new content. These breakthroughs have led to a new generation of storytelling systems that enable open-ended conversation and create pedagogically beneficial text and images [81, 132]. In these ways, GenAI models expand the ability of teachers to facilitate open-ended discussions with underserved youth

[82]. This new technology opens up new possibilities for developing literacy education platforms built on top of the new technology capabilities - such as creating artificial intelligence agents to facilitate human-to-AI communication [27]. To integrate such emerging technology effectively into educational settings, it is critical to understand the practical challenges and needs faced in classrooms and the complexities of teachers' roles, deeply rooted in the reality of teaching and learning [82, 113].

### 3.3.3 Learning at Scale

In examining the history of learning at scale, the instructional approaches of Massive Open Online Courses (MOOCs) and Intelligent Tutoring Systems (ITS) could offer valuable insights into how GenAI can be integrated into education. These models provide contrasting paradigms of educational delivery, each with its own set of strengths and limitations. Understanding these approaches could illuminate strategies for integrating generative AI into educational settings more effectively. MOOCs are primarily driven by instructor-led methodologies, whereas ITS are dependent on algorithmic guidance. Each approach has its unique advantages and limitations. Starting with MOOCs, which are instructor-guided platforms, often struggle with personalized engagement due to their large-scale, standardized format. This can lead to a lack of individualized attention and feedback, essential components in the learning process. Similarly, ITS, one notable limitation is in the area of algorithmic customization. There is a lack of flexibility in this framework, which often just differentiates between 'right' and 'wrong' answers. Such binary assessments fail to capture the nuances and complexities of learning, potentially oversimplifying educational experiences and processes. These limitations are crucial to consider as we explore the potential of generative AI in educational settings. It's important to explore how these GenAI systems can be designed to not only address these limitations but also to improve adaptive, interactive, and personalized learning experiences.

To gain insights into how generative AI can enhance personalized learning platforms in education, it is essential to review the history of large-scale learning platforms. Notably, platforms like Massive Open Online Courses (MOOCs) have significant limitations in providing personalized experiences. This is largely due to their vast scale and reliance on standardized curricula, which do not easily accommodate individual learning needs. Another example of an Intelligent Tutoring System (ITS) has its limited capabilities to provide a flexible learning experience because of the binary assessment method that evaluates exclusively right or wrong answers which oversimplifies learning experiences. It is imperative to find a balance between the two different ways of learning (i.e., who guides the learners) with a new lens of the current technical capabilities with pedagogical approaches [144] with GenAI.

## 3.4 Co-Design Logic Model with Teachers

### 3.4.1 Participants

We conducted co-design sessions with six primary school teachers in the United States. The co-design sessions took place individually via video conferencing due to geographical distances, with an average length of approximately one hour and a half for three days (i.e., once a week) between July and September 2023. Teachers were recruited through snowball recruitment as part of researchers' networks, with the only criteria for eligibility being that they were either current or former K-12 teachers. The teachers were classroom teachers from 1st to 8th grades, and their experience averaged 15.8 years (min=2.3 years, max=31 years).

On the first day of the co-design session, we sought to elicit teachers' current practices, their struggles, and their motivations when they teach writing to their students. Afterward, in the second session, we introduced GenAI platforms (i.e., features and functionalities). We asked about their experiences and opinions about adapting it in educational settings

specific to writing activities with their students. Then, co-developing logic models and sharing feedback and suggestions were discussed. Our study was approved by the authors' institutions' institutional review boards (IRBs).

### 3.4.2 Data Sources and Analysis

We analyzed the qualitative data with a deductive approach from the co-design sessions. First, the interview data was automatically transcribed (Otter.ai) using the original audio and aligned with the transcript. The transcript was transferred to qualitative data analysis software (Atlas.ai) for the first round of coding. We conducted a deductive approach to analyze interview data (Azungah., 2018) from the logic model structure to identify context, inputs, output, and outcomes. We read the transcripts and identified relevant themes of the text. We categorized codes into four high-level themes (i.e., context, inputs, outputs, and outcomes) and 27 codes under each theme (see Figure 1). The construction of categories from the data was driven by the research objective to understand what features and functions the educators desire to implement into AI-based writing tutoring platforms. After the categories were consolidated, the researchers considered how the categories relate to models of writing education [10, 17], and theoretical work on writing support [8]. Although not explicitly considered in the creation of the categories, this analysis constituted a further analysis of the relationship between the categories, as well as the significance and implications of what the results imply.

## 3.5 Result from the Co-Design sessions

From the findings, we identified the context and users of the AI-based writing tutoring platform for emergent writers and readers. The results revealed input stations in which

students do, and what features are necessary for the platform to encompass: providing playful literacy activities, text chats, selecting culturally responsive imagery, writing models, feedback, and translanguaging [54]. Drawing from teachers' interviews and synthesizing evidence-based research literature, we demonstrated effective strategies in literacy instruction for youths as outputs of the logic model, which is about what students learn as follows, genre-specific writing strategies, fluent in spelling, typing, word processing, lexical diversity, and verbal reasoning [73]. Lastly, we drew desired outcomes from interview data based on teachers' perspectives, such as improving students' self-efficacy and motivation in writing, as articulated by one teacher:

*"I think that as a writing teacher, a lot of us want the students to feel curious, feel confident, feel motivated, feel activated as writers are encouraged. I think a lot of writing teachers will say 'I can work with the student whose writing is weak', but getting them to start, getting them to continue and finish, and then getting them to revise. Those are the psychological hurdles, right?"*

A teacher's goal is to support a student's writing activities by providing writers' tools through the platform, which we consider to be an input of the platform. What we provide as features includes providing age-appropriate and immediate feedback to students and scaffolding mentor text that is helpful to students in developing their writing.

*"We can ask students, do you want to use the thesaurus? Not only are you letting them have a choice, but you're reminding them of writers' tools. I think what you want your writers to do is to be able to improve their writing, by improving their workflow choices, their structure, sentence skills, organization, and so paragraphing."*

Teachers are intrigued by the potential of generative AI (LLM) to offer customized mentor texts, catering specifically to students' needs, including varying vocabulary levels and phonics.

*"It still is very time-consuming to make sure that you're finding good and mentor texts that fit with what you're teaching to make sure you're hitting direct discrete skills that are real. Using mentor text is really a lot of teachers' work. That is one of the ways that we can use AI – to help us develop appropriate levels of writing models for each student in different states."*

Our findings indicated teachers identified that the integration of gamification and culturally relevant imagery generation via Text-to-Image AI art generator (TTL) serve as effective strategies to augment engagement and motivation within writing activities.

*"And it seems like it's a game that you can play with the AI. Like, you write the description of the alien? you give the prompt, AI gives you the image."*

Teachers highlighted the potential of utilizing AI art generators (TTL) to produce culturally relevant images, thereby amplifying student engagement during story creation.

*"Who's your character? Are they white? Are they black? You know, so that now they can have something that represents them? Right? If they're a black kid, they want to see a story about a black kid, they don't want to see a white kid with blond hair all the time."*

Consequently, we identified the following logic models: Context/Users (struggling writers and low-level readers), Inputs/Activities (playful literacy activities, text chats, culturally responsive imagery selection, writing models, feedback, translanguaging, and writing strategies), Outputs (increased amount of time spent in writing, word count, genre-specific writing strategies), and Outcomes as fluent in spelling, typing, word processing, lexical diversity, syntax, and verbal reasoning, ultimately increase engagement and motivation in writing (see Figure 1). These can be facilitated via the development of AI-based writing tutoring platforms through conversational design, chatbot systems, and customizable editing stations (Appendix ).

## 3.6 Design and Development

As detailed features as inputs from the logic model that we co-designed, we developed an AI-based Writing Tutoring Platform (AWTP) based on features and functionalities within the input section in the logic model. One key feature is that students can generate mentor texts as models for their writing, a popular strategy in writing instruction [29]. We developed AI agents that deliver real-time feedback on the quality of students' writing. These agents not only identify grammatical and spelling errors but also offer constructive suggestions to enhance the students' writing skills.

In addition to that, we tailored the lesson modules to offer genre-specific writing lessons (i.e., narrative writing, and opinion writing). In each writing module, we aim to provide instructions about what genre (i.e., fiction and non-fiction) and topics associated with the genres into writing lessons. Such as, narrative writing modules include writing stories about fairytales, science fiction, diary, and memoir. Opinion writing modules provide topic selection associated with the pros and cons of certain issues (i.e., Should Animals Be Kept in Zoos?, Homework: Helpful or Harmful?).

Teachers have noted that students, particularly those who are emergent writers, often struggle to initiate their writing from scratch. To address this, we designed an AI agent to facilitate brainstorming sessions. In these sessions, students can explore topics they are passionate about, such as school uniforms, homework, or favorite books. These topics are specifically chosen to engage students' interests and are relevant to their everyday experiences.

We intended to design the AI agent to address teachers' concerns that their students struggle both with starting their writing and effectively completing their pieces. Teachers reported that students often write just a few sentences and consider their work finished. To overcome this, the AI agent is designed to help initiate the writing process and provide ongoing guidance through the introduction, body, and conclusion stages. For instance, in narrative

writing, students have a chance to create important components in stories, (i.e., characters, settings, conflict, and climax) to map out their stories. Furthermore, we have implemented an image generation feature to help students to visualize their stories with a goal of enhancing students' engagement in story creation. Opinion writing module, AI agents guide students' thought processes by initiating conversations about students to state their opinions, asking them to provide reasons to support their opinions with examples. In conclusion, AI agents are designed to ask students to summarize their opinions and reasons.

Recognizing the potential benefits for English Language Learning and settings involving English as a Second Language for emergent readers and writers, as well as immigrant families, we aimed to incorporate translation features that will convert students' input into English. We are currently refining the interface design to enhance its usability. Additionally, teachers have suggested integrating Text-to-Speech (TTS) and Speech-to-Text (STT) functionalities to support young students transitioning from reading to learning to read, emergent readers and writers, and to uphold principles of inclusive and universal design. Teachers have also recommended the inclusion of a dictionary or thesaurus to assist emergent readers in understanding content generated by the AI, providing them with additional support in comprehending new vocabulary and concepts.

### 3.6.1 System Overview

Following the design strategies outlined previously, we developed an (AWTP). This platform integrates an AI-enabled chatbot, a mechanism for processing user-defined conversational protocols, and a flexible canvas element to display their story map as graphic organizers to help students structure their stories. We aim to provide students with an engaging learning experience that mirrors the interactive dynamics of a writing activity.

**Story Lesson Processor**

The central system module processes a sequence of story creation lesson steps outlined in a JSON file. These steps, which include predefined actions, are then passed on to the Chatbot module for execution. Each lesson step is assigned a unique identifier for internal reference purposes, thereby enabling the articulation of the processing sequence with a modifiable execution order. The innovation of the system is attributed to the assortment of pre-defined modular actions for any given lesson step. Each action type is associated with a distinct service, encompassing but not limited to, completion services provided by Large Language Models (LLMs), text-to-image services, image-to-text services, text-comparison services, and calculation services, among others. Given the modular design of the actions, lesson authors possess the full autonomy to integrate any combination of action types and quantities to construct an effective lesson step. This flexibility facilitates the generation of tailored and instantaneous knowledge content for users.

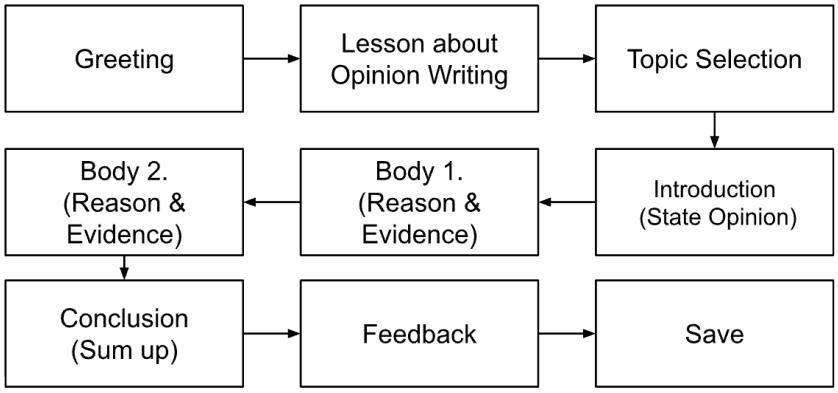


Figure 3.1: Interactive processes of the AWTP opinion writing module.

## 3.7 Usability Study

To understand the applicability, affordance, and limitations of the AI-powered writing application in learning settings, we conducted a usability study with thirteen students ages 8

Table 3.1: Writing workshop schedule

| | Day 1 | Day 2 |
|---|---|---|
| Topic | Should Animals Be Kept in Zoos? | Homework: Helpful or Harmful? |
| Group A | Writing with AWTP | Writing without AWTP |
| Group B | Writing without AWTP | Writing with AWTP |

to 10. We conducted within-subject experiments, splitting the students into two groups. On the first day, one group utilized AWTP while the other group engaged in traditional writing exercises without AI assistance (writing on a Google doc) (see Table 3.1). The groups swapped AI assistance on their writing activity the following day. This alternation procedure was designed to minimize any potential order effects from the use of AI-assisted writing.

### 3.7.1 Procedure

We conducted a two-day writing workshop with thirteen students in an out-of-school setting in Southern California, United States. Students were asked to conduct a pre-interview about their writing efficacy and AI experience, then we divided the group of the students who were writing opinion essays with and without AWTP, randomly assigned each day. On the first day, group A students were asked to use the AI-assisted writing platform to write an opinion essay with a topic (i.e., Should Animals Be Kept in Zoos?). The next day, students write a different topic (topic b (i.e., Homework: Helpful or Harmful?) without an AI platform on a Google doc. The next day, the groups exchanged tasks, alternating between writing activities with and without AI assistance. Lastly, we conducted a post-interview on the second day of the workshop.

### 3.7.2 Participants

Thirteen participants aged eight through ten were recruited through snowball sampling from our institution via advertisements on mailing lists. The average age of the participants was 9. 2 years, and 8 of the 13 participants were girls. One-third of the participants predominantly spoke English at home, and the rest spoke another language, mainly Chinese and Korean. The majority of participants (n = 10) identified as Asian, while two identified as White, and one as more than one ethnicity. All of the participants' parents had a Bachelor's degree or higher. All participants reported that they had experience using AI applications (i.e., Alexa, Google Home) before, while two used them monthly, and another six used them more frequently (i.e., weekly or daily).

### 3.7.3 Data Sources and Analysis

We screen-recorded students' writing activities and collected students' writing outputs as text files. To understand the applicability of the 'output: direct result/data from the system experience' from the logic model, we measured students' time spent in writing, total word count, and lexical diversity to elicit affordances of the application in students' writing efficacy. We recorded students' utterances using voice recorders as well as saved student-AI utterances with AI-assisted writing platforms to understand the usability of the platform.

To analyze students' writing output, we utilized Text Inspector, a web-based text analysis tool [164] to measure lexical diversity. The Text Inspector gives the Measure of Textual Lexical Diversity (MTLD) which is a measure of the range and variety of vocabulary used in a piece of writing or speech. It reflects the number of different words used, and often the uniqueness of those words, within a given text or across texts. Higher lexical diversity indicates a broader vocabulary and can enhance the richness and depth of communication, making it an important factor in evaluating language proficiency and complexity in both

| | MTLD | | TWC | | Time Spent (mins) | |
|---|---|---|---|---|---|---|
| | P-value: 0.004** | | P-value: 0.03* | | P-value: 0.001** | |
| | M | SD | M | SD | M | SD |
| Without AWTP | 64.90 | 12.65 | 162.62 | 60.59 | 26.15 | 13.31 |
| With AWTP | 79.83 | 9.57 | 210.77 | 79.79 | 36.77 | 16.05 |

spoken and written language [120]. Lexical diversity is important in writing because it can enhance the readability and interest of a text, prevent repetitiveness, and demonstrate the writer's linguistic competence. It's particularly relevant in creative writing, academic writing, and other contexts where a rich vocabulary is beneficial.

### 3.7.4 Result from the usability study

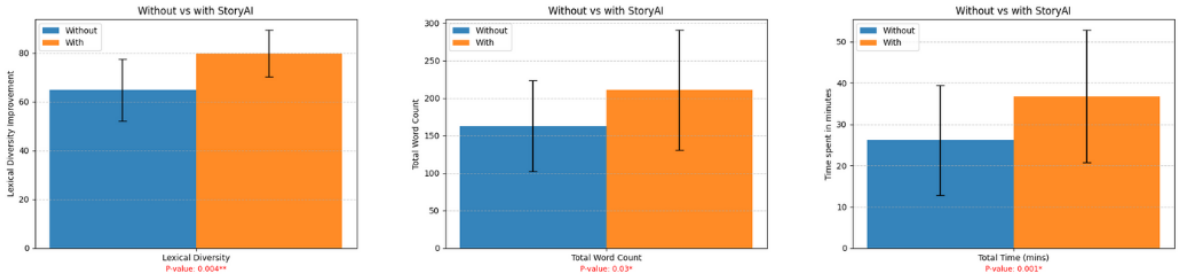


Figure 3.2: Result from the writing workshop

The quantitative analysis results revealed that students stayed in writing longer when they utilized AWTP for their writing, the average writing time for students without AWTP was 26.15 minutes and they stayed 10.62 minutes longer when they wrote with AWTP. Additionally, students tended to write more (total word counts) with AWTP as they wrote an average of 162.62 words but they wrote 210.77 words when they wrote with AWTP which indicated the AWTP's potential to promote writing engagement [78]. We also noticed students' lexical diversity [119] in their writing has increased.

## 3.8 Feedback Study

In addition to the usability study with students, we conducted feedback interviews with fifteen educators. The majority of the educators (n=11) are experienced K-12 classroom teachers. Each interview took an average of 25.7 minutes. Before the interview, teachers conducted a short survey, they were asked to try the AWTP platform with a temporary URL that we created for the survey and asked for their opinions on the applicability of AWTP in educational settings.

### 3.8.1 Participants

Fifteen teachers participated in the study, whom we recruited via a mailing list and snowball sampling through our network in and out of the institution and local communities. Ten participants identified as female, and five participants identified as male. The median years of experience is over 16 years (see Table 3.2) and the median age group is 45 to 54 years old. Nine participants were located in California, five were from Washington, and one was from North Carolina, United States. All participants provided informed consent to participate in compliance with our institution's IRB. Participants also provided their availability, which we used to select the workshop dates and times.

### 3.8.2 Data Sources and Analysis

We first conducted a survey on understanding the usability and applicability of AWTP in educational settings, in the survey, teachers were asked to try the AWTP platform of which the URL is publicly available during the period. The survey included demographics (i.e., occupation, years of experience, ages, gender, and location) and their opinions about AWTP on the Likert scale (i.e., How likely would you be to recommend this tool to your friends and

colleagues? Do you think this tool would help your students develop overall as writers?).

Educators who agreed to further interviews were recruited following the survey. One-on-one interviews with teachers for approximately 30 minutes the interview questions included their general in-depth opinions about AWTP. The interview protocols included as follows,

- What initial reactions do you have to our current design? Did you have a positive experience as a user of the app? Why or why not?
- Do you think this tool would help your students develop overall as writers? Please explain why.
- Would you have any concerns about using the tool with students? If so, please describe.
- What features would you recommend adding to the tool, and why?

All interviews were video-recorded, automatically transcribed through Zoom, and manually revised to correct errors afterward. We used a general inductive approach for analyzing qualitative data. The first author open-coded the transcripts to identify patterns in the dataset. The research team discussed and identified themes. The final codebook contained three parent codes and eight child codes for the feedback study. By discussing the categories repeatedly, the researchers reduced overlap and created low-level categories that were shared. A high-level category was then formed from these low-level categories. During the later meetings, a third researcher gave further insights into the data as the researchers repeatedly met and iterated to construct these categories. Participant recruitment continued until saturation of themes was reached during this analysis, which was conducted concurrently with the interviews. A low-level category was annotated for each relevant text segment at the end of the analysis process.

Table 3.2: Participants' information for the feedback study

| Alias | Occupation | Years of experience | Location | Gender |
|---|---|---|---|---|
| T1 | Edutech Specialist | 1-3 years | California, USA | Man |
| T2 | Ed tech | 8-10 years | California, USA | Woman |
| T3 | Spanish Teacher | More than 16 years | California, USA | Woman |
| T4 | Teacher | More than 16 years | California, USA | Man |
| T5 | Edu tech | More than 16 years | Washington, USA | Woman |
| T6 | English teacher | More than 16 years | California, USA | Woman |
| T7 | Teacher | More than 16 years | California, USA | Woman |
| T8 | Teacher | More than 16 years | California, USA | Woman |
| T9 | Teacher | 8-10 years | North Carolina, USA | Woman |
| T10 | Teacher | More than 16 years | California, USA | Man |
| T11 | Teacher | More than 16 years | Washington, USA | Man |
| T12 | Teacher | More than 16 years | California, USA | Woman |
| T13 | Teacher | More than 16 years | Washington, USA | Woman |
| T14 | Teacher | More than 16 years | Washington, USA | Woman |
| T15 | Lecturer | 3-5 years | Washington, USA | Man |

### 3.8.3 Results from the feedback study

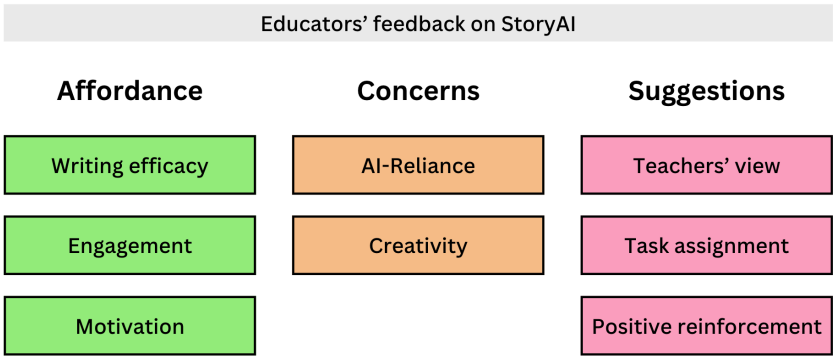


Figure 3.3: Findings from the educators' feedback

From the qualitative analysis, we elicit overarching themes into three categories: affordance in writing, concerns leveraging AWTP in learning environments, and design suggestions from teachers' perspectives in improving AWTP to meet the needs of their unique teaching environment. In each category, there are two or three subcategories that identify what impacts the high-level categories. This study aims to better understand educators' perspectives on AI-based writing tutoring platforms.

### Affordance

We categorized the themes of affordance about AWTP's effectiveness, engagement, and motivation in writing for students. Teachers have highlighted the potential benefits of AWTP, noting its ability to help students get started on their writing assignments, maintain focus, and enhance motivation by improving their self-efficacy in writing. One of the teachers mentioned that AWPT can help students reduce writing anxiety specifically for struggling writers. As T2 stated,

*"Writing can be so intimidating and I think this would help struggling students become more motivated and less intimidated."*

Also, T10 emphasized that AWTP is engaging for students and offers valuable opportunities for collaborative learning by stating,

*"AWTP was easy and fun to use. The prompts keep them going. They would enjoy writing and sharing their work with classmates, a tool to promote collaborative learning among peers."*

This would help to motivate students to write more and encourage them to collaborate with their peers.

### Concerns

The qualitative data uncovered educators' concerns regarding the use of automated generative artificial intelligence functions in writing. First of all, teachers expressed worries about the possibility of students becoming overly reliant on AI assistance. As T13 remarked,

*"My biggest concern is that my students depend on and rely on this app most of the time, so they cannot write a story without StoryAI."*

Similarly, educators expressed uncertainty over students' independent learning experiences, as some students might just passively engage in writing experience by accepting AI-suggested text and not engage actively in writing activities. This could lead to a lack of creativity and an inability to express their ideas and opinions. In addition, AI-generated text may not reflect a student's unique voice and style, resulting in a less authentic writing experience.

**Suggestion**

Educators also provided some design suggestions. First of all, educators proposed the development of adaptive interfaces that enable students to modulate the degree of AI assistance they receive. Additionally, they recommended the implementation of teachers' monitoring systems that allow teachers to evaluate student engagement. Such systems would enable educators to determine whether students are passively relying on AI support or actively engaging with the learning material. T13 said,

*"Can teachers see the writing process to verify that students were actively engaged rather than passively accepting the AI's help? Maybe a choice on how much AI can be used for the process."*

This would ensure that students are actively working on the writing process rather than passively relying on AI. Furthermore, educators reported having a teacher's view to create classes and assign specific tasks for their students. T7 commented,

*"It would be useful to be able to create classes and assign tasks to students. I would also like to be able to see how often the students rely on AI assistance."*

These features would enable teachers to restrict the level of AI assistance students can utilize, thereby providing teachers with more control over the educational process and the ability to tailor tasks to meet the unique learning needs of each student. Additionally, it would

provide teachers with valuable insight into their students' writing performance and progress.

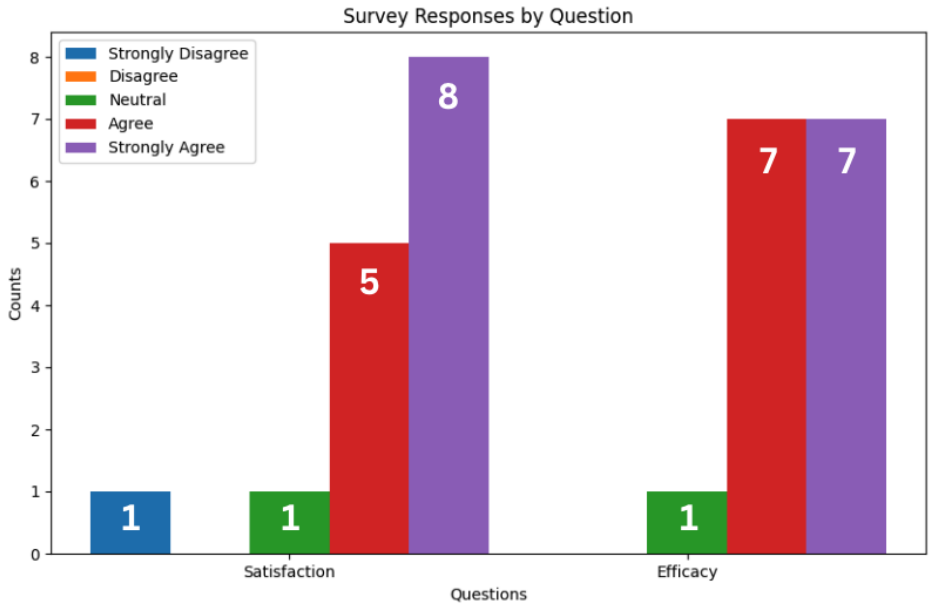


Figure 3.4: Result from the educators' survey

Further feedback survey results indicated that teachers provided optimistic perceptions on using AWTP for their students' writing. The majority of teachers (13/15, 87%) showed their desire to recommend the system to their colleagues, and most of them (14/15, 93%) think AWTP helps students develop writing abilities.

## 3.9 Discussion

The study represents an initial step in navigating the emergence of AI tools that can be integrated into learning experiences by constructing a logic model for the design and development of AI-powered writing education platforms tailored for young learners. In order for AI-driven educational technology to be effective, it needs to resonate with educators' needs, concerns, and aspirations. We developed the application based on the logic model we elicited with teachers, as a result of our collaborative efforts with K-12 educators and synthesized academic literature. Our approach underscores the importance of involving educators from

the beginning, fostering an informed perspective on the design of learning technologies.

Our findings emphasize that while Generative AI offers exciting new possibilities, its adaptation in K-12 educational settings should be approached thoughtfully. Key aspects include ensuring cultural relevance, amplifying student engagement, and nurturing learners' motivation and self-efficacy in writing. The teachers' perspectives illuminated the value of integrating gamified elements and culturally responsive imagery, which not only elevates the writing experience but also validates and respects students' identities.

Furthermore, the discussion around mentor texts, especially the potential for AI to offer tailored writing examples, highlights the transformative potential of AI. This could alter the teacher's role from that of a content provider to a facilitator, offering students more autonomy in their learning journey and equipping them with tools that cater to their specific needs. However, as with any innovative technology, challenges remain. The concerns about authenticity, agency, and the potential for bias or misinformation underscore the need for rigorous testing, regular updates, and perhaps most importantly, ongoing dialogue with educators and stakeholders.

### 3.9.1 Design Recommendations

The usability study with students and the feedback study with teachers have provided valuable design insights, which we have summarized into two categories. The first set of insights aims to enhance student experiences within AWTP usage. The second set focuses on suggestions elicited that aim at enabling educators to more effectively utilize the platform via design approaches, specifically tailored to their distinct teaching contexts based on their feedback.

### For Students

From the study, we noticed several areas for improvement through the design iteration of the platform to enhance students-AI collaboration on writing and learning. The first approach is to facilitate writing conferences for learners to be able to share their writing pieces with others, comment on each others' essays, and promote collaboration among peers. The second is to scaffold learning experiences in terms of both interaction techniques (i.e., conversation agent vs. chatbot), and modalities (i.e., videos, animations, and text), as well as age-appropriate learning materials (i.e., topics and vocabularies), through implementing various learning support, including, video lessons, text-to-speech, and speech to text for voice-based conversation; dictionary, thesaurus, and translation. Lastly, we suggest contextually appropriate and personalized students-AI dialogue and interaction, through conversation design strategies with teachers where designing AI agents to build a rapport with students (i.e., ask students' feelings) for better social Interaction; and personalization. Students were also given several suggestions where they would like to have more culturally salient user experiences in terms of image generation, as well as teachers pointed out the importance of tutorials on how students and teachers can navigate the platform necessary to effectively utilize the platform.

### For Educators

The study offered valuable design insights for educators, enabling them to enhance the platform. It allows educators to tailor their instruction, assign specific tasks, and provide step-by-step guidance to meet their students' individual learning needs. To meet the needs of educators' unique needs and mental models, we recommend developers and researchers conduct co-design sessions to prototype the interfaces with teachers. This includes card-sorting activities where teachers have the agency to categorize ideas based on their understanding of

the interface (i.e., subjects, grades, and topics). This gives us insight into the way teachers think about the interface, allowing us to design it in a way that meets their needs.

The study revealed that teachers have the desire to keep track of students' writing progress and learning outcomes. This can be provided through learning analytics systems, which can be facilitated as a dashboard for teachers to easily understand students' learning outcomes and patterns. The dashboard and learning analytics strategies could be co-designed with teachers to identify specific learning competencies they wish to monitor and determine the most effective visualizations for teachers to comprehend. This can help teachers to better identify and address learning gaps, as well as target areas for focused instruction. Additionally, learning analytics can help teachers to individualize instruction and tailor it to the specific needs of each student. This collaborative process will ensure that teachers can use the dashboard to make informed decisions and strategies to improve their teaching practices.

**Limitation**

This study has several limitations, including first, the small sample sizes for co-design sessions with six teachers as these insights might bring less diversified perspectives in terms of diverse subjects and grade levels. The same as for the within-subject experiment with thirteen students which is not sufficient and could be generalized to a larger group of students. Despite efforts to mitigate order effects—where the sequence of conditions influences results—by dividing participants into two distinct groups, and attempts to reduce carryover effects—where the influence of prior conditions affects subsequent conditions—by assigning two different writing topics, eliminating the possibility of such effects remains challenging.

Another limitation concerns the measurement metrics utilized, specifically total word count and time spent writing. Although these metrics are grounded in the literature regarding the efficacy of writing interventions for struggling writers [83, 76], it is difficult to assert that they

definitively represent the primary competencies for assessing writing efficacy. Furthermore, from the usability study, we have not yet investigated the screen recordings or AI conversation logs, which could offer deeper insights into how students interact with the system and apply StoryAI to enhance their writing strategies. Analyzing these data would reveal the extent of students' reliance on AI assistance and their independent learning experiences, which are crucial for understanding their engagement in writing.

Additionally, the feedback study involving fifteen educators presents unique limitations. The feedback, based on the educators' opinions regarding the potential applicability of the platform in K-12 classroom settings, could have been more informative if conducted with teachers who have experience using StoryAI with their students. Such firsthand experiences would provide a more comprehensive understanding of the platform's real-world applicability in educational settings.

## 3.10 Conclusion

This series of studies began with the co-design of a logic model in collaboration with educators, aimed at designing and developing a prototype for an AI-based Writing Tutoring Platform (AWTP). The research progressed by conducting a usability study with students and a feedback study with educators. The results of the study suggest that AWTP has the potential to improve students' writing outcomes, specifically engagement and efficacy by improving students' time spent in writing, total word count, as well as lexical diversity in their writing product. The study also provides valuable insights into the usability of AWTP and the expectations of educators. As educators pointed out, the affordances of AWTP are to promote writing efficacy, engagement, and motivation for their students. They also pointed out several constraints, including over-reliance on AI and skepticism regarding students' capacity for independent learning when using the AWTP for writing. Thereby,

educators suggested incorporating educators' views into the system so that they can assign their unique tasks through classroom setup, as well as provide moderate AI assistance and active engagement in learning experiences for their students. Additionally, educators provided positive opinions about the current AWTP design as they are willing to recommend the platform to their colleagues, and they think AWTP helps students enhance their writing skills.

Consequently, we provided recommendations for designers and developers to enhance AI-based learning systems to include pedagogical approaches, such as writing conferences and scaffolding for vocabulary enhancement, as well as the implementation of additional lessons. Furthermore, we emphasized the integration of social elements, such as fostering rapport with AI (through conversation design), and inclusive design features like translation, text-to-speech (TTS), and speech-to-text (STT) services to facilitate personalization and adaptation to the unique needs of each student. Additionally, these suggestions enable teachers to customize their lessons and effectively manage their classes through collaboration.

In conclusion, the study's contribution lies in not only providing a pathway for the design of an AI-based writing tutoring platform but also in emphasizing the value of a collaborative approach in the development procedure with stakeholders in education (i.e., teachers). By bridging the expertise of developers, researchers, and AI specialists with the lived experiences of educators, it is possible to ensure the development of innovative educational platforms that are both cutting-edge and grounded in real-world educational needs. By constructing a logic model that outlines the blueprint of an AI application—from design and development to evaluation with learners and educators—we provide empirical evidence of its potential utility in education. This evidence can inform the future development of AI-based Writing Tutoring Platforms (AWTP). Our long-term aim is to develop a system that creates a feedback loop between developers, researchers, and educators, allowing for continuous refinement and improvements of the AWTP.

# Chapter 4

# StoryAI: Generative AI-powered story-authoring platform for young learners

## 4.1 Study Abstract

The ability to tell a story is correlated with success in writing. However, it is hard to facilitate one-on-one storytelling activities in an educational setting. We conducted a formative study to understand students' (ages 8-12) capabilities and needs in leveraging the Generative AI platforms (i.e., ChatGPT, Stable Diffusion) in writing via visual story creation activity. With the design insights from the workshop (i.e., we designed and developed StoryAI, a text-based conversational agent that facilitates storytelling activities that guide and promote story writing for children. We conducted a user study with thirty students. We found StoryAI's efficacy in students' perception of writing competencies (i.e., planning, translating, and revising) as well as AI literacy (i.e., perception, confidence, and motivation). Based

on qualitative data, we categorized the main themes into affordance (i.e., motivation, engagement, and learning), efficacy (i.e., planning, translating, and revising), and perceptions (i.e., ownership, constraints, and suggestions). At the end, we discuss design suggestions and research topics that need to be further examined to inform the future development of AI-enabled story-writing platforms for educational purposes.

## 4.2 Introduction

Storytelling is an effective pedagogical strategy that can be incorporated into lessons to increase students' competencies across various disciplines [137, 142, 123, 84]. Due to the interrelated nature of the processes involved in reading and writing and social interactive elements in storytelling among listeners and tellers, storytelling has advanced participation and engagement in learning experiences [51, 201, 191]. Digital Storytelling platforms offer various advantages to promote learning not only literacy (i.e., reading and writing), but also abilities to convey meaning to others, convey information, and express their thoughts via interactive format with various media such as image, video, and text [95]. However, due to the individualized nature of storytelling activity involving a listener and a teller pair, it is hard to facilitate one-on-one interactions for all learners in educational settings. To benefit more students from storytelling activities, text-based, and voice-based conversational agents have been utilized [191, 93, 108, 127] that can engage in meaningful conversations, as it does not require the presence of another person.

The text-based conversational agent, a chatbot, has been studied and utilized to improve personalized and adaptive learning across a broad range of educational contexts, including literacy [57, 201, 189], mathematics [185, 199], and system thinking [127]; it has demonstrated effectiveness not only in learning itself but in advancing motivation by reducing anxiety [157, 110]. However, most previous works involved rule-based chatbot systems, which were

limited in their scope of interaction [158, 94]. The rapid progress in generative AI (GenAI) technologies, exemplified by innovations like ChatGPT, and Stable Diffusion, has opened up unprecedented opportunities for dynamic conversations and creative activities, capturing significant interest within the field of HCI and educational research [106, 82]. Furthermore, the advent of Large Language Model (LLM) technology has significantly enhanced the capabilities of these AI agents, enabling them to deliver responses that are remarkably human-like and replete with accurate information [82]. Historically, chatbot applications have utilized straightforward, textual interfaces to afford users the ability to retrieve information, engage with services, or partake in entertainment via online messaging platforms [158]. More recent efforts in the field have aimed to expand the utility of chatbots, employing them to assist individuals facing social and communicative challenges, or to encourage users to pursue domain-specific learning objectives, such as language acquisition and storytelling [57]. The preference for chatbots as a medium of informational interaction stems from their ease of use, naturalness, and intuitive design.

However, due to the complex and dynamic nature of the educational environment, especially in interactions between teachers and students, the integration of advanced technologies into educational environments demands careful consideration. This is essential to ensure their safety and effectiveness in improving both teaching and learning processes [144]. With increasing concerns about academic integrity and the potential over-reliance on AI within educational settings [82], it is critical to consider strategic design considerations of GenAI-powered learning applications. Innovative strategies are necessary to address the challenges posed by GenAI in education while maximizing its potential benefits, such as personalization and adaptive learning [96].

From this study, our objective is to identify the essential considerations for designing and developing GenAI-powered story-authoring platforms tailored to children who are emergent readers and writers [20]. Additionally, we focus on the iterative design of text-based conver-

sational agents to assess the efficacy and applicability of the platform in writing education. To elicit design strategies that can be better suited for students ages 8 to 12, we conducted a formative study involving a creative visual story writing activity with the preexisting GenAI platforms like ChatGPT and Stable diffusion [12, 53]. With the findings from the formative study (i.e., lack of context, independent writing learning opportunity, and autonomy) we determine design strategies to mitigate these weaknesses and integrate them into our system design and development of StoryAI (see Figure 4.1). We developed StoryAI to function as a storytelling collaborator that guides students through the narrative writing process—from ideation and planning to structuring their stories. StoryAI incorporates text-based chatbot technologies alongside an open-ended, customizable canvas, empowering students to create and personalize their visual stories. Furthermore, it enables students to save their projects locally, facilitating ongoing access and revision.

The study is guided by the following research questions: How can we design an AI-powered story-authoring platform for students that is accessible, safe, and effective? How do students interact with and benefit from the features of the GenAI-powered story authoring platform in their learning and writing processes? What are the affordances and limitations of StoryAI on creative writing?

To answer these questions, we conducted a usability study with the high-fidelity platform, StoryAI, at one of the non-profit community centers on the West Coast, California, USA with thirty students ages 6 to 14 for four days, three hours per day. We administered pre- and post-surveys (i.e., first and the last day) aimed at evaluating students' writing efficacy and motivation, as well as their understanding of AI literacy [114]. Additionally, we screen-recorded students' writing activities and collected their writing outputs in a text file. To assess students' perceptions of ownership over their written outputs and their opinions on their experiences with StoryAI, we conducted individual semi-structured interviews with students following the completion of their writing activities.

From the usability study, we sought to identify the applicability and constraints of StoryAI in narrative writing for youths. From the pre-and post survey, we identified students' writing efficacy has increased in three areas, planning, translating, and revising. Not only the writing efficiency, students' AI literacy has enhanced, such as their perception of AI, confidence in using AI for their tasks, and motivation to learn more about AI. From the qualitative analysis of the semi-structured interviews of students' opinions about their experience of StoryAI, we uncovered three themes for their writing, 1) affordance in motivation, engagement, and learning, 2) efficacy in planning, translating, and revising, and 3) perception about ownership, constraint, and suggestions. To be more specific, StoryAI helps students to get started on writing easily and students enjoy writing with StoryAI. The AI agent allows students to get started on writing and students feel attentive to the AI agent which indicates its affordances in social interaction between AI and a child.

Additionally, StoryAI's agent supports students' writing processes and guides them along the writing journey, educating them on genre-specific writing strategies and grammar. However, students also expressed a sense of diminished ownership over their writing outputs, as they often incorporate ideas from AI-suggested texts. They expressed a desire for greater control over their writing process. Thereby, students provided design suggestions like color-coded students' writing output over AI-generated text, as well as the ability to customize the level of AI assistance in their projects.

The contribution of the work, first, began by informing designing and developing strategies of the system in a collaborative process of designing a GenAI story authoring platform for writing activities with stakeholders in education through formative studies. Second, we provide strategies for how to design and incorporate AI interaction with open-ended creation to promote students' autonomy in writing and integrate storytelling activity into craft narrative writing for students. Also, we provide insights into restrictions associated with AI reliance for students for independent learning experiences. The research will offer empirical

evidence on the applicability of GenAI in educational settings, contributing to a comprehensive understanding of its benefits and limitations. This evidence can guide the design and implementation of GenAI in educational contexts. Ultimately, the study will provide insights into how advanced technology, like generative AI, can be aligned with educational goals and age-appropriate learning experiences, bridging the gap between technological advancements and educational needs.

## 4.3 Literature Review

### 4.3.1 Co-writing systems in HCI research

Co-writing systems in HCI research have been focused on mostly adult users and their perception and interactive style to understand the efficacy and applicability of the AI-assisted co-writing systems [68, 104, 198]. In terms of the Cognitive Process Theory of Writing, there are three interconnected processes in writing starting from planning, translating, and reviewing [62]. Past AI-writing supporting tools mainly focused either the planning part (i.e., inspiration and ideation) [spark, ], or the reviewing stage (i.e., grammar, spelling, style checking tools like Grammarly, Wordtune) and focused more on the writing technique for professionals and adults writers less focus on storytelling perspectives, fewer attempts to integrate storytelling activity to writing in a learning platform.

Biermann et al.[22] studied writers' perception of AI-co-writing systems towards writers' autonomy and ownership and examined what motivated story writers' AI support to elicit AI companion design. They found barriers and motivations to leveraging AI support, one of the barriers was the writer's desire to take control of their writing and fulfill their integrity and ownership over their writing. Another barrier is when writers distrust AI capacities. Story writers are willing to utilize and motivate AI support when they focus more on productivity

and cognitive offload when they are not confident enough in their writing abilities. Biermann et al. [22] pointed out the importance of writing responsibilities between writers and the AI companion to balance autonomy and control in a human-centered AI system.

Gero et al. [68] presented a system called "sparks" that generated inspirational sentences for scientific writing and tested the efficacy of the system with 13 STEM graduate students. Gero et al. [68] found three major use cases of sparks to inspire writers: translation (i.e., help writers to translate ideas into a sentence), and perspective (i.e., understand different perspectives). They found users who initiated more inspiration and perspective when using Sparks tend to be more independent writers, but those who take translation take advantage of AI interaction more often. Schmitt & Buscheck [153] studied CharacterChat which they designed a chatbot that initiates conversation over creating a character of a story, they found the chatbot style interaction helps writers to get inspiration on the attributes of characters for their stories in the early stage of the character development.

Yuan et al. [198] studied how people leverage AI-powered co-writing systems for creative writing and found effectiveness in reducing writer's block and promoting creative processes. Yuan et al. [198] suggest enabling user customization to improve engagement and provide users autonomy in their writing which benefits from having both a combination of pre-build control to open-ended customization – which is also been highlighted in Biermann et al. [22]'s study of which writers' sense of ownership is important to motivate their AI companion interaction. So our platform StoryAI focuses on designing a balanced approach to AI guidance with open-ended customization.

Lee et al. [104] studied LLMs' applicability in understanding interaction techniques among writers and AI by building a CoAuthor that captures x y z. The integration of LLMs in practical applications has highlighted the need for robust datasets that can provide insights into their interactive capability The creation of the CoAuthor dataset involved structured writing sessions where writers interacted with different configurations of GPT-3, performing

both creative and argumentative writing tasks. The sessions were meticulously recorded, allowing researchers to analyze the nuances of AI-human collaboration. The dataset, along with a user interface for replaying these sessions, is made publicly available, offering a valuable tool for the HCI community to study LLMs' interactions. The analysis of the CoAuthor dataset revealed several key insights into GPT-3.5's capabilities as a collaborative partner in writing. The dataset demonstrated that GPT-3.5 could contribute effectively to various stages of the writing process, including generating ideas, formulating arguments, and revising text. The study also highlighted the contextual and sometimes subjective interpretation of AI performance, emphasizing the need for nuanced metrics to evaluate collaboration quality. Our approach is to design an AI companion function not just as a source of inspiration (ideation tools) but as a translation from ideas to text.

### 4.3.2 Leverage LLM to enhance teaching and learning

The purpose of this section is to discuss how LLM and GAI have been integrated into educational research. AI-powered educational tools have made significant advancements in two main areas: STEM (Science, Technology, Engineering, and Mathematics) and language education, which encompasses literacy, reading comprehension, and writing skills. In STEM education, LLMs contribute to creating dynamic learning environments that can simulate real-world problems and provide instant feedback, thereby fostering a deeper understanding of complex concepts. Meanwhile, in language learning, these models support the development of reading and writing skills through personalized learning experiences that adapt to the individual's proficiency level and learning pace.

**STEM + AI**

As one of the example projects in STEM, Xu et al. [185] leverage voice-based conversational agents to promote mathematical concept conversation among parents and children. From the need-finding interviews with family groups, Xu et al. [185] 2023 designed MathKingdom that teaches children math concepts (i.e., measurement, sequence, patterns) through the multimodal voice-guided games that they found CA is effective in children's engagement as well as enhanced children's mathematical languages and abilities, as well as promote child-parent conversation about math. Also, Xu et al. [185] conducted a study to investigate the effects of conversational AI integration into narrative science programming for children. The study centered around interactive videos where the main character, powered by a conversational agent, interacts directly with the viewers by asking questions and flexibly responding to their inputs. The interactive video enabled real-time conversation between the children and the video character, aiming to mimic a social interaction.

Assessments were conducted immediately after viewing to evaluate the children's understanding of the science content presented in the episode. The findings from Xu et al. [185] demonstrated that children who engaged with the conversational AI character showed significantly better performance in the science assessments compared to their peers who viewed the non-interactive version. Moreover, the study noted that parents often played a supportive role during the interactive sessions, enhancing the learning experience.

Dietz et al. [51] introduced StoryCoder, a voice-based AI application that guides young learners in computational thinking concepts through storytelling so that they can build computing and literacy skills together. The research conducted by Dietz et al. [51] highlights the effectiveness of using storytelling in a voice-guided app to teach CT to young children. The project was initiated to bridge the gap between literacy skills with computational thinking education for young learners who are not literate enough to read and understand textural

and numeric concepts in CT. With a multi-day user study with children, the authors found the effectiveness of introducing target computing concepts (i.e., sequences, loops, events, and variables) via story game in StoryCoder. Results from the evaluation study indicated that children were able to successfully navigate the app and effectively learn the targeted CT concepts. Notably, children demonstrated above-chance performance on the CT concept recognition task after engaging with the app, suggesting that the integration of storytelling into CT education can lead to meaningful learning outcomes. The findings underscore the potential of storytelling as a powerful tool for teaching complex concepts to young learners. By embedding CT education within a narrative context, StoryCoder not only makes learning more accessible but also more engaging for children. This approach aligns with educational practices that emphasize active and meaningful learning experiences.

**Language Arts & Literacy + AI**

The project by Xu et al. [189] developed and tested "Rosita Reads With My Family," a conversational agent coupled with an e-book, designed to facilitate parent-child interaction during reading sessions. The agent was specifically tailored to accommodate the linguistic and cultural characteristics of Latinx Spanish-English bilingual families. The development process involved iterative design and testing to ensure cultural and linguistic appropriateness. The researchers conducted a user study with Spanish-English bilingual families to evaluate the effectiveness of the conversational agent in engaging children and promoting parental involvement. The study assessed verbal interactions and the quality of engagement during the reading sessions. Results from the user study indicated that the bilingual conversational agent successfully engaged children verbally and enhanced parent-child interactions during shared reading sessions. Children demonstrated active participation and improved language use in both Spanish and English, while parents reported greater ease and interest in participating in bilingual reading activities.

Jeon 2024 [94] explored the affordance and efficacy of employing a custom chatbot for domain-specific educational experiences in the classroom for English as Foreign Language (EFL) students. They seek to understand students' perception and affordance of the chatbot for their learning experiences. They found three main areas of affordance of leveraging chatbots in class: pedagogical, technological, and social. They highlight the interaction between students and the chatbot agent as the interlocutor enhances active engagement in language learning experiences. Social affordances develop students' positive attitudes towards their learning as students feel less pressure from their mistakes and social stigma from being judged by others. Even with the advantages, some of the students preferred human partners over chatbots (especially for those who are already good at English), because they want to play English games with human partners more openly and explore language learning experiences. Also, some students showed frustration when the chatbot didn't understand their utterances due to their accents.

### 4.3.3 A review of the current Gen-AI story-authoring platforms

Numerous applications are being developed that leverage generative AI to support creative expression and story creation [5, 2, 3, 8, 6]. While many applications cater to adult users, we focused on choosing apps designed specifically for children and families. Some apps facilitate the creation of storybooks for children [1, 4, 7]. However, these applications have not been thoroughly examined for their impact on learning and literacy, leaving it uncertain whether their primary function is educational or purely for entertainment.

As the purpose of the study is to design a GenAI-powered story authoring platform for young learners for their literacy development, we intend to analyze the selected applications to assess their effectiveness in supporting the cognitive processes involved in writing from planning, translating, and revising [62] as well as affordance in receptive language (i.e.,

listening and reading) and expressive language (i.e., writing and speaking) acquisition. We selected ten web applications based on the usage of generative AI in its systems for story creation for children and we leveraged search terms, *AI and storytelling*, *AI storytelling*, and *generative AI story applications for children*.

Our analysis revealed that all of the applications provide users with the option of generating stories based on a simple description (i.e., describe your story). The primary features of these web applications include AI-generated quick stories, which promote the creation of personalized and customized narratives for users. Some applications further enable users to set parameters such as age groups, image styles, and narrative perspectives (first-person or third-person) as part of the planning process, such as StoryNest.ai [6] and CreateBookAI [41]. For instance, Scarlett Panda [4] provides a story page that includes sections for users to input a brief description of the story, details about the main character's friends, the moral of the story, the language used, the target audience's age group, and the illustrated style. CreateBookAI [41] allows users to type the title of a story, the name of the character, and their companion. They also have a section for users to pick a setting of the story (i.e., Middle Age, Pirate, and Future Times), and moral of the story (i.e., Courage, Tolerance, Justice, and Charity), and pages of the story. Once users pick and explain them in a short passage, the AI generates a whole story, then users can save the illustrated stories in a PDF file.

The majority of the applications allow users to listen to a story that is AI-generated and customize fonts, text, and images [5]. Most applications support receptive language acquisition, such as reading and listening even have translation features. For instance, StoryWizard [7] generates a complete story then users can choose to listen to or read the story, with an integrated dictionary feature available by double-clicking on unfamiliar words. Additionally, users can edit stories by changing names or text and regenerating images. However, these features are not scaffolded to aid children effectively. The application lacks feedback or guidance to help children comprehend the language they are reading or listening to, and it does

not offer support for teaching children how to construct their own stories.

Even though these applications have several ways to facilitate planning the story, and idea support (i.e., character creation) [1] they are missing the 'translation' part of the writing process which is critical in their ability to translate ideas into sentences. These applications are one-time one-shot story generation, which enables users to type a sentence about a theme of the story and the system generates a whole story [8]. The majority of applications fail to implement scaffolding in story creation and the cognitive process of writing as they don't allow users to brainstorm, translate with multiple trials, or revise. There is a notable deficiency in the available methods for users to edit and revise their work, as the system lacks feedback mechanisms or guidance on the revision process. Consequently, it falls upon the users to determine how to effectively utilize the system for their revision needs.

## 4.4 Method

### 4.4.1 Formative Study

To elicit design implications and understand how students leverage the current version of generative AI chatbots and image generators on the market (i.e., ChatGPT, Stable Diffusion), we conducted a formative study with the platforms with twelve students ages eight to twelve. Twelve students were asked to create a visual story using ChatGPT and Stable Diffusion and work on their final output in Google Docs. We focused on our observation notes, screen recordings, and writing outputs to elicit design implications from students' writing strategies leveraging the platforms.

### 4.4.2 Participants

We conducted a workshop with children ages 8 to 12 (2nd and 6th graders) to understand students' strategies and struggles when interacting with existing chatbots and text-to-image generators. Participants were recruited using a snowball sampling method from our researcher's network (mailing list and contacts). Parents (n=12) completed a screening survey before attending the workshop to make sure they were 18 years or older and lived with children between the ages of 8 and 12. According to parent reports, the mean age of the student participants was 9.8 years old, and five (5/12, 42%) were girls. Eleven children (11/12, 92%) were identified as Asian American, four children (4/12, 33%) spoke only English at home, and the remainder were bilingual (6/12, 50%) or spoke English as a second language (2/12, 17%). All children possessed sufficient oral English proficiency for daily conversation. It was the first time the students had used the GenAI-LLM chatbot and Text-to-Image generators (TTL), while (7/12, 58%) of parents reported having already used them.

### 4.4.3 Procedures

The 2-hour, 1-day workshop was conducted in a community center 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) [53](Stable Diffusion, 2023) and a chatbot powered by LLM (i.e., ChatGPT) (see Figure x). During the writing project, we sought to understand the students' strategies and their interactions with the system through observation by taking a field note and voice recording youths' verbal expressions and semi-structured interviews [146]. 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.

### 4.4.4 Formative Study Findings

The formative study identified three primary areas for improvement. First, students struggled to initiate their writing activities because ChatGPT did not provide sufficient context, requiring formal guidance on how to start. Second, students often used AI-generated outputs directly, which contradicted the objective of learning to write through generative AI. Third, given their varying levels of digital fluency, students found it challenging to navigate between websites and transfer their work from one platform to another, such as copying from ChatGPT to a Google Doc.

#### Lack of context

We noticed students had a hard time actively initiating the conversation due to the lack of context that was provided by the platform at the beginning of the interaction, which often resulted in incomplete or superficial user experiences (i.e., typing unrelated, no-context words like Hahaha). This issue was compounded by the limited capacity for users to initiate writing to accommodate their unique needs or to improve the platform's overall accessibility and user-friendliness for the age group. The constrained nature of these dialogues highlights the necessity for more intuitive and context-aware AI systems that can proactively guide

students in navigating the complexities of their writing tasks, ultimately fostering a more engaging and productive learning environment.

**Lack of independent writing learning opportunity**

The writing process involves several stages, from brainstorming, and planning to drafting and revising [62]. However, the current design of ChatGPT interfaces does not adequately support this iterative process of writing. The interfaces rely heavily on users' ability to initiate 'prompting' to guide the system through their writing process, which can be particularly challenging for students aged eight to twelve—a critical period for literacy development. Given that writing involves more than a one-off generation of text including planning, drafting, and revising, and needs a certain degree of cognitive development [20], it's imperative to design the platform to promote and scaffold the processes of writing for students. Such a redesign should enable a more collaborative interaction between the user and the chatbot, facilitating the translation of ideas onto a digital canvas and supporting iterative revisions and reflections on both students and generated content.

**Lack of autonomy**

From our observations, we indicated that students faced challenges when navigating between different websites—ChatGPT to Stable Diffusion and Stable Diffusion to Google doc— which highlighted the necessity for a platform to be redesigned so that it allows youths to easily navigate across functions and features. Additionally, we noted a distinct lack of customizable options, which constrained students' ability to personalize their writing experience. This made it difficult for students to express themselves fully within the digital space, making it difficult to communicate their ideas effectively. Furthermore, the limited customization options made it challenging for students to create a unique writing output. This lack of

flexibility prevented students from using the platform in ways that best suited their individual ideas and needs, posing a significant barrier to enhancing student agency in their writing.

## 4.5 Design Strategies

Our formative study identified several challenges faced by students during their initial interactions with the system. Firstly, students struggled to get started with their tasks without a clear context, leading to confusion. Secondly, there was a noticeable dependency on AI-generated text, which was contrary to the independent learning objectives that teachers and parents advocated for. Thirdly, there is a lack of customization features, as well as cognitive overload, was apparent as students had to navigate between multiple platforms like ChatGPT (for their writing), Stable Diffusion (for their visual image generation), and Google Docs to compile their written and visual stories. To overcome these obstacles and enhance the learning experience, we established the following design objectives: 1) predefined clear structure of AI-assisted learning paths to actively support students' writing processes, as such, we designed the lesson instructions (i.e., conversation) with teachers on how AI agents can carry over the lessons effectively. 2) minimize students' reliance on AI for content generation to foster independent thinking by designing AI's role as a listener and students as teller, allocating AI's role of a writing partner and collaborator (instead of an assistant or helper) who can scaffold, encourage, and provide feedback on students' creative writing processes, and 3) add customization stations for open-ended personalization by students seamlessly integrate AI assistance with options.

For the instruction, we intended to provide a way of promoting the process of writing from planning (i.e., topic ideation, character development, image generation, and graphic organizers), translating (i.e., check writing, AI-reliance checker, feedback), and revising (i.e., editor, reviewing stage, and saving) [62]. Our overarching goal is to find a balanced strategy to

integrate instructor-led and algorithmic-led learning experiences to integrate the flexibility of GPT's open-ended conversational capabilities with the structured guidance of instructor-led learning experiences. Therefore, we implemented a chatbot-style interface as its primary mode of interaction, empowering storytelling activity that leads to creative writing practices by providing both guidance from a chatbot and an open-ended creative canvas.

Drawing from the empirical evidence collected during the writing workshop and synthesizing relevant literature, we developed effective design strategies. Chi & Wylie (2014) developed the ICAP framework, where the researcher examines cognitive aspects of engagement in learning processes (i.e., interactive, constructive, active, and passive) [36]. The researcher argued that interactive activities such as dialogues provided opportunities to develop constructive knowledge through the mutual exchanges of thoughts and ideas, and thus can enhance learning [48, 143]. Therefore, we intend to design text-based conversational agents that can carry over the active interaction with students. Chi & Wylie 2014 also pointed out that generating and correcting concept mapping (i.e., graphic organizer of knowledge) is a mode of constructive engagement which is an activity that constructs knowledge acquisition [85, 125]. The authors indicated that the Interactive mode of constructing concepts is even better in learning individually and creating maps [47].

Further, Xu and Warschauer's literature on the content analysis of voice-based apps [193] identified key aspects that influence children's learning and engagement with voice-based applications. The authors stressed the importance of facilitating interactive learning activities and clarifying goals for young learners to improve active participation and reduce cognitive load by allowing youths to focus on current tasks [19, 183].

## 4.6 STORY-AI: Bringing storytelling (Chatbot) and narrative writing (Canvas)

Following the design strategies outlined previously, we developed StoryAI, a platform for story authoring and learning. This platform integrates an AI-enabled chatbot, a mechanism for processing user-defined conversational protocols, and a flexible canvas element. It aims to provide students with an engaging learning experience that mirrors the interactive dynamics of a storytelling activity. Simultaneously, it offers educators the ability to facilitate adaptable and progressive instructional sequences that the chatbot can execute.

StoryAI is a combination of a text-based conversational agent with an open-creation canvas– the AI agent has a smooth conversation with students and guides students in story creation and students can work on their story on the right side of the screen. The agent provided choices of selection for students (i.e., genre, topic, characters, and setting). The AI agent guides students on story structure from introduction, conflict, rising action, climax, falling action, and resolution. The AI agent guides students on story structure from introduction, conflict, rising action, climax, falling action, and resolution. The functionalities and interaction were designed to foster students' agency in their writing so that they can take control of their writing project and have a greater sense of ownership by encouraging students to write their own stories with the help of an AI agent as a companion rather than an assistant. We provided story genre selection and topics, and the agent carried over the step-by-step guidance on developing writing projects, idea generation, and adding detail (Table 4.2.).

### 4.6.1 Features

#### Agent Features

To facilitate students' interaction with AI agents for story-building, we implement chatbots as the primary mode of interaction. It is essential to develop a conversation design that effectively supports writing. This approach is grounded in a synthesis of the literature on writing education [73], cognitive engagement of learning [36] and the cognitive process of writing [62]. We designed the conversation with prompts strategies to be effective in learning interaction by scaffolding and feedback. Based on these cognitive theories of writing processes and standardized curriculums of writing education, our AI agent is designed to promote the following stages of writing starting with planning (i.e., idea generation, character development), translating their ideas to sentences, then reviewing and revise their story with the strategies of conversational prompts.

To support scaffolding in students' learning, we facilitated elaborative feedback where AI agents follow up and rephrase students' responses as an extension of students' language, to promote dialogic interactions as well as cognitive engagement [97]. We leverage a balanced approach of rule-based and LLM-based approach in designing the conversation through the process of planning, translating, and reviewing. Such as genre choices and topic selections were designed as rule-based we designed them with experienced writing educators. However, idea suggestions for story components (i.e., settings, characters' characteristics) were LLM-based as AI agents provide options for students to pick and expand their ideas. In the planning stage, we intend to design the agent to co-construct a concept map (see Figure 4.1) for a story (i.e., idea, character, setting, conflict) as the interactive mode is the most effective in learning. Collectively generating and correcting concept maps results better than individual creation of the maps [47].

**Student-AI interaction**

| Interaction | Prompts | Example Output by AI agent |
|---|---|---|
| Guidance | Encouragement | "Alright, first, do you want to pick a genre? Now, can you think of an idea to write a story based on a Fairy Tale?" |
| Suggestions | Created "I need help" button for students to trigger ideas from AI agent | That's ok, here are some ideas for you... |
| Feedback | Prompt strategies: | 1. Clarity: The story has a clear beginning, middle, and end, making it easy to follow.<br>2. Quality: The plot is creative and engaging, with unexpected twists<br>3. Grammar: Check for punctuation errors and improve sentence structure for better flow.<br>4. Word choices: Use more descriptive words and varied vocabulary to enhance the imagery.<br>5. Sentence structure: Vary sentence lengths and structures to make the writing more dynamic and engaging for the reader. |

Table 4.1: An example of a conversation between a student and the AI agent.

**Platform feature**

StoryAI has three essential components, first, the main page is where students can pick various writing genres (i.e., narrative writing, opinions writing). In each module, students can further explore the genre of each writing (i.e., fiction and non-fiction) and different topics.

In this design, alongside the incorporation of a text-based conversational agent interface, we introduce a canvas-based mode of interaction to provide students with graphic organizers to

plan their writing. These can help them structure their thoughts and reasons systematically. This approach is designed to facilitate the documentation and organization of information, the formulation of ideas, and the creation of a learning artifact emanating from engaging dialogues with the chatbot. The canvas component is crafted to maximize creative freedom for users, enabling them to input text, manipulate images and videos via drag-and-drop functionalities, perform basic drawing tasks, and develop rudimentary animations. We posit that the process of ideation, augmented through learning, constitutes an invaluable practice. Numerous platforms have been developed based on this foundational concept to enhance ideation and creativity, rooted in the essential notion of a visual interface within the sphere of human cognition. Visual systems, especially those aimed at fostering creativity (including, but not limited to, mind mapping tools, drawing applications, chart utilities, and word processing software), universally incorporate the canvas element as the base, underscoring its essential contribution to facilitating visual creativity and organizational capabilities.

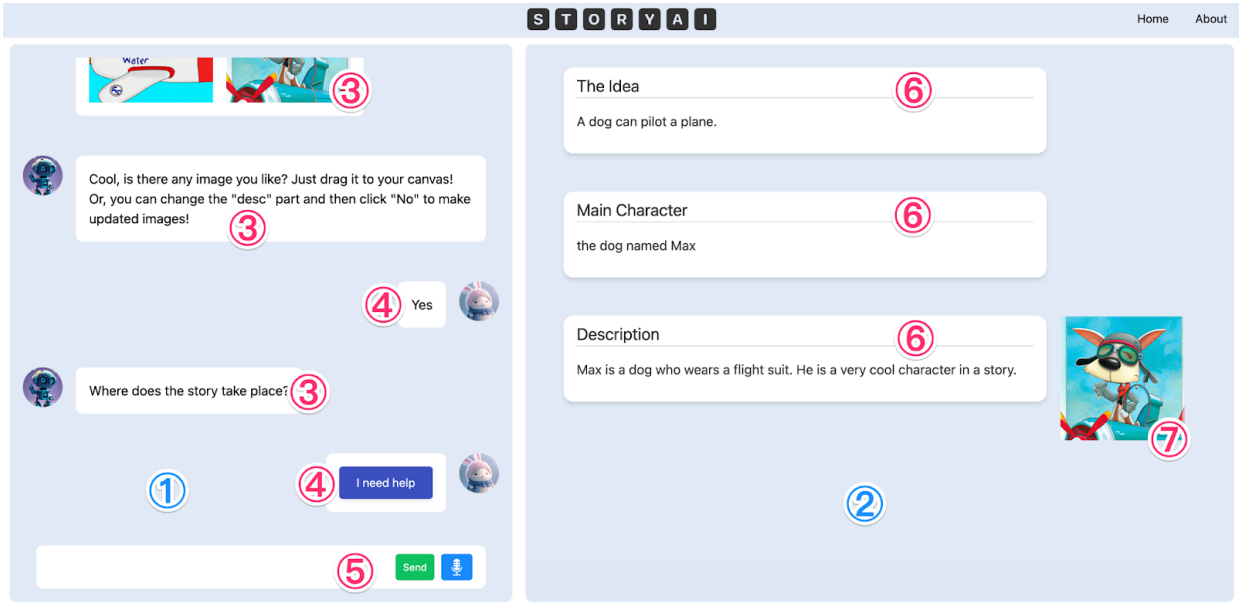


Figure 4.1: (1) Chatbot panel, (2) Canvas panel, (3) Chatbot response, (4) option buttons, (5) input message field, (6) writing section (7) multimedia items.

Within the interface, there are two panels: the Chatbot panel (1) and the Canvas panel (2). The chatbot station serves as the platform where students engage in conversations with AI agents. These AI agents operate based on a combination of rule-based and large language

model (LLM)–based interactions to provide instruction. The canvas side of the interface interacts with the chatbot to display a conceptual map of the story, allowing students to customize, review, and revise their writing. These panels do not interact with each other implicitly, except when user-initiated actions occur, such as dragging an image from the Chatbot panel and dropping it onto the Canvas panel. This design enables users to effectively manage their focus, allowing them to either engage in an inspirational conversation with the chatbot or develop their thoughts on the canvas and translate their ideas to text. Based on the actions executed, the Chatbot will respond (3) in diverse formats and associated multimedia content, including text, images, audio, video, buttons, and segments of web pages. Whenever a user interacts with the Chatbot by entering text or clicking on a button, the response is logged in the conversation thread (4) for future reference.

The Chatbot interface features a minimally designed input field (5) for users to type their chat messages. This field is specifically intended for brief messages, hence its compact size. Additionally, we have integrated a Voice-to-Text feature, complemented by a microphone button next to the text input field, allowing users to dictate their messages without needing to type.

In the Canvas panel, users have the freedom to relocate any information blocks, including text (6) and image (7) blocks, and can modify the content of these blocks using the provided configurations. Our goal is to enable users to construct their ideation effortlessly. Ultimately, this canvas will serve as a learning artifact or outcome, which could be a presentation, poster, essay, or storybook.

To carry over the smooth conversation between AI and students, we implemented Text-To-Speech (TTS) and Speech-to-Text (STT) where students can not only type in but also they can speak their stories to the AI agent, who can then respond accordingly. Students also can use different languages other than English, the platform automatically translates the input into English to carry over their stories in English. To support scaffolding, we implemented

the "help" button where students can ask questions so an AI agent can answer that question openly. Students can also generate images according to their story where they can regenerate and customize the images in the canvas.

| | **Chatbot** | **Canvas** |
|---|---|---|
| Activity | Storytelling | Narrative writing |
| Learning | Brainstorming, Ideation, Planning | Organizing, Structure |
| Modality | Communication | Publication |
| Interactivity | On the fly; Dynamic | Craft, Revising |
| Medium | Immediacy | Permanence |

Table 4.2: Interactive processes of the application.

### Development Framework

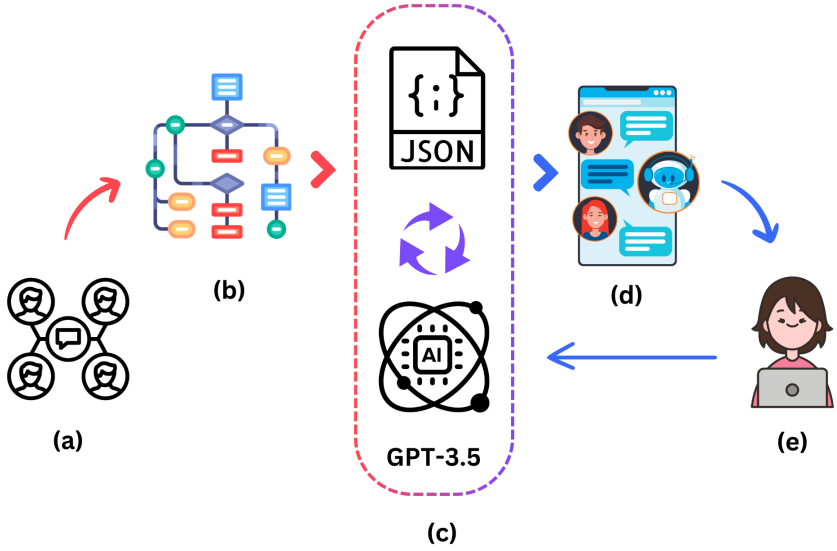


Figure 4.2: (a) Lesson plan designed by educator (b) translate the steps of lessons into logic flow ⓒ translate the wireframe to the chain of prompt in JSON and send step by step to GPT 3.5 (d) vue.js front-end view appears on the screen as a chatbot (e) students interact with the system and send their emergent utterance and writing to the GPT 3.5.

### 4.6.2 System development

We implemented StoryAI using OpenAI API (GPT-3.5) in Vue.js. By developing a web application, we made it easier for educators to integrate it into their lesson plans, and for students to access it anywhere without downloading any software. StoryAI helped us evaluate, test, and iterate the design more easily. The application is publicly accessible at (On-going iteratively designing and developing the platform): https://story-ai.co/ (see Figure 4.1)

We developed StoryAI, through co-designing the curriculums and lesson plans with teachers. These step-by-step lesson plans are converted into a flow chart for the AI instructor, which is translated into a chain of prompts in the JSON file that carries over the carefully crafted prompt engineering strategies with teachers that send it to GPT3.5 step by step. Utilizing the OpenAI API (GPT-3.5), we enabled the AI to use these prompts in a dynamic and interactive mode, ensuring a balanced approach, blending AI's dynamic conversation with the focused, educational direction of human educators.

The aim is to create AI agents capable of executing lesson plans (i.e., creative story writing) effectively. On the other side, we intend to allow students to openly iterate and edit their stories along with the student-AI conversation. The interface, designed using Vue.js, presents a text-based conversational agent that students can interact with to enhance their writing skills. We then implemented the open-ended creative canvas on the right side of the screen to allow students to openly edit and iterate their visual story (see Figure 4.1).

Figure 4.3 is the snapshot of how we translate from the lesson plans conversation design flowchart to a JSON file with a chain of prompt engineering strategies. This approach enabled us to integrate the flexibility of GPT's open-ended conversational capabilities with the structured guidance of instructor-led learning experiences, ensuring a balanced approach, and blending AI's dynamic conversation with the focused, educational direction of human

educators. Consequently, our system adopts a chatbot-style interface as its primary mode of interaction, facilitating the dissemination of information to users and guiding users through the flow of system operations with AI-enhanced support.

### System Overview

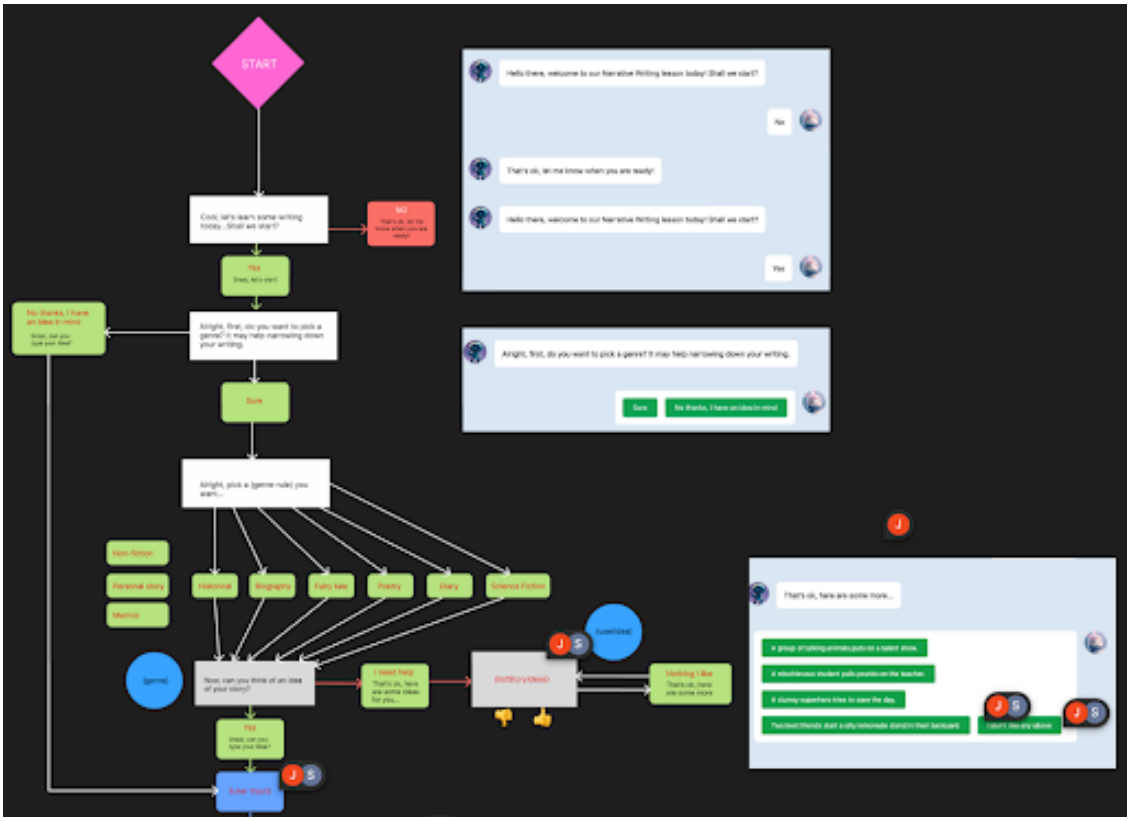


Figure 4.3: System Flowchart Process

### Story Lesson Processor

The central system module processes a sequence of story creation lesson steps outlined in a JSON file. These steps, which include predefined actions, are then passed on to the Chatbot module for execution. Each lesson step is assigned a unique identifier for internal reference purposes, thereby enabling the articulation of the processing sequence with a modifiable execution order. The innovation of the system is attributed to the assortment of pre-defined

modular actions for any given lesson step. Each action type is associated with a distinct service, encompassing but not limited to, completion services provided by Large Language Models (LLMs), text-to-image services, image-to-text services, text-comparison services, and calculation services, among others. Given the modular design of the actions, lesson authors possess the full autonomy to integrate any combination of action types and quantities to construct an effective lesson step. This flexibility facilitates the generation of tailored and instantaneous knowledge content for users.

## 4.7 Usability Study

To understand the applicability, affordance, and limitations of StoryAI in learning settings, we conducted a user test study with students (n= 30). We conducted pre-and post surveys on their writing efficacy, motivation, and AI knowledge. We also interviewed them after they used StoryAI. Additionally, we screen-recorded students' writing processes and collected their writing outputs as well as student-AI utterances. From the data collection, we aimed to uncover the following research questions,

- In what ways does StoryAI help students enhance their writing efficacy and motivation?
- Would writing with StoryAI help students develop AI literacy?
- In what ways do students make use of StoryAI to support their writing?
- How do students perceive StoryAI AI agents for their writing?

| Day 1 | Day 2 | Day 3 | Day 4 |
|---|---|---|---|
| Pre-survey | Writing with StoryAI | Writing with StoryAI | Post-survey |
| Lesson<br>What is AI? | Interview | Interview | Discussion |

### 4.7.1 Setting and Participants

The research was conducted at a non-profit community center in Southern California during the spring of 2024. This center is well-integrated with the local school districts, predominantly serving Latinx communities. We engaged a diverse group of 30 participants, aged 6 to 14 years (mean = 9.73, SD = 2.26), including 16 females and 14 males, all of whom identified as Hispanic or Latinx. The study spanned four days within a single week, organized into three age-specific groups. Each group participated in a one-hour session daily.

### 4.7.2 Study Procedure

On the first day, we built a rapport at the beginning of the session, introduced each other, and shared what they liked and disliked the most for the first 15 minutes of the session. Then students started to conduct a pre-survey test, it took around 25 mins to finish. Then students were taught an AI lesson that went over a simple concept of artificial intelligence (i.e., what is AI, AI examples). The second day of the workshop encompassed using the StoryAI platform for narrative writing. Students began their writing module and started writing without any guidance other than StoryAI guidance itself. The writing activity lasted around 35 minutes on average for students and the rest 15 minutes took place for individual semi-structured interviews, two of the researchers at the site recorded the interviews. On the third day, the same process was carried out. On the fourth day, students took post-survey questionnaires. It took an average of 30 minutes for students to finish. We took the remaining 20 minutes to discuss what they liked and disliked about using StoryAI for their writing and how they can make StoryAI better. Students who finish their writing and the

rest draw 'how StoryAI might look like' on a paper.

### 4.7.3 Data sources and analysis

#### Pre- and Post Surveys

We collected and analyzed pre- and post-survey from 19 participants among the total 30 participants in the workshop, because 11 students conducted either pre or post. We couldn't compare the differences. Consequently, the 11 students were excluded from the analysis of the pre-and post-survey. The pre-and post-tests were designed to examine writing motivation that focuses on planning (coming up with ideas), translating (expressing ideas to a sentence), and revising (editing their text). The other area that we focused on was AI literacy [114], specifically, students' perception (what is AI), confidence (confidence in using AI), and motivation (I want to use AI for my writing) and asked students to write with our image prompt to examine their fluency and flexibility in writing. Aside from the pre-post survey, we also asked students' efficacy in writing (do you like writing?) and current knowledge and experience on AI (do you have any experience with AI daily?) in the post-survey, In post-test, we added questions regarding their experience on writing with StoryAI platform for StoryAI efficacy StoryAI experience and their opinion focus on their sense of ownership over their writing output, enjoyment of using StoryAI for their writing, ease of use, collaboration, and their satisfaction over their writing output (see Appendix. A).

Video Observations, Interviews, and Writing outputs We screen-recorded students' monitors to capture students' writing processes and strategies. Also, we collected students' writing output in text files as well as AI-child interaction dialogue. To understand students' overall experience and their opinions on StoryAI, we conducted a 1:1 interview with each child after they finished their writing. Two of the researchers took turns to take semi-structured interviews (see Appendix A.1 for interview protocols). The interview questions were mostly

open questions, asking students to explain their experience, what they liked and disliked StoryAI, and if they liked their writing outputs and their perception of AI's role overwriting (i.e., ownership) and their suggestions. First, the interview data was transcribed using an automatic transcription program (Otter.ai) that retained the original audio and aligned it with the transcript. After a thorough review of the transcript, we transferred the transcript to qualitative data analysis software (Atlas.ai) to do the first round of open coding (Salda, 2013). We conducted an inductive approach to analyze interview data (Thomas., 2006). Following the inductive approach, two researchers independently read the transcripts and identified relevant themes of the text. 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 GenAI; how their values and motivations differ) into each theme. To reduce the 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 and continued until a saturation of themes was found. We organized our results around the main theme of the advantages and challenges of using StoryAI for narrative writing activity, which emerged from this coding. We categorized codes into three high-level themes (i.e., affordance, efficacy, perception, and suggestion). The analysis contained nine mid-level themes (i.e., motivation, engagement, learning, planning, translating, revising, ownership, limitation) and 24 codes under each theme.

## 4.8 Result

We carried out quantitative and qualitative data analysis to understand the applicability and efficacy of StoryAI for students' writing. The following results explore how StoryAI affords students' learning and areas to be improved for further design implications.

| Key Competencies | Gain | | | Pre-test | | Post-test | |
|---|---|---|---|---|---|---|---|
| | M | t | P | M | SD | M | SD |
| **Writing habit** | | | | | | | |
| Planning | 1.26 | **3.73 | 0.001 | 2.32 | 1.06 | 3.58 | 1.02 |
| Translating | 1.1 | **3.54 | 0.002 | 2.53 | 0.90 | 3.63 | 1.01 |
| Revising | 0.95 | **3.16 | 0.005 | 2.26 | 0.87 | 3.21 | 0.98 |
| **AI literacy** | | | | | | | |
| Perception | 0.62 | **2.25 | 0.04 | 3.32 | 0.75 | 3.94 | 0.94 |
| Confidence | 0.89 | **3.03 | 0.007 | 2.95 | 1.08 | 3.84 | 0.69 |
| Motivation | 1.25 | **3.83 | 0.001 | 2.58 | 1.02 | 3.83 | 0.99 |

### 4.8.1 Pre- Survey

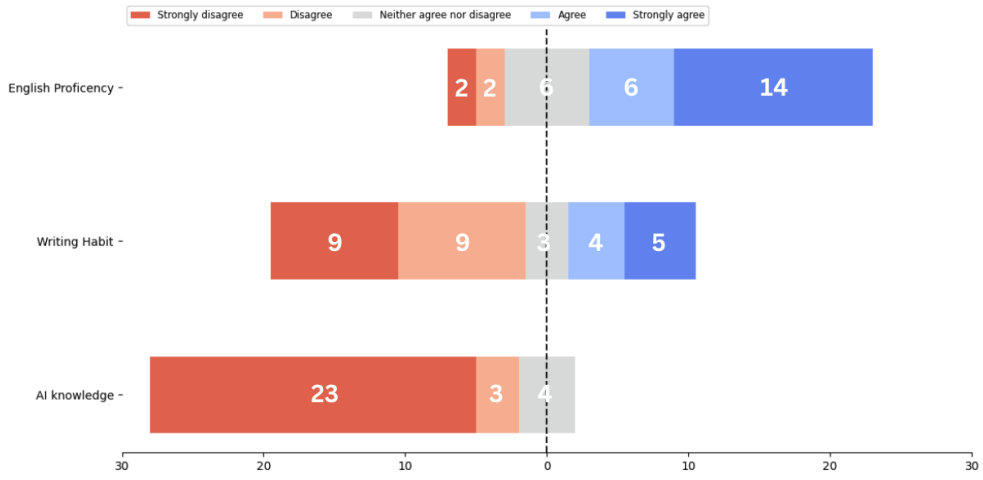


Figure 4.4: Pre-survey results on the English language proficiency, writing habit, and AI knowledge.

In order to understand students' language proficiency as well as writing habits, we asked questions about their English proficiency rates and how much they enjoy writing stories and essays on a Likert scale in the pre-survey. The survey results revealed that the majority of the students (20/30, 67%) have decent English proficiency, and the rest indicated themselves as not fluent in English (4/30, 13%). For their writing habits (i.e, do you like writing essays and stories), eighteen students (18/30, 60%) indicated they dislike writing at home and at school, and only nine (9/30, 30%) students reported they enjoy writing which means, the majority of students have a lack of writing efficacy and motivation (see Figure 4.4).

In addition to the prior writing efficacy, we sought to identify students' prior knowledge of artificial intelligence before the session, thus we presented questions about their current usage of AI technology in their daily activities. The result revealed that all of the students (30/30, 100%) did not have prior AI experiences or AI knowledge and experiences.

**RQ. Would the StoryAI experience help improve students' perception of their writing efficiency?**

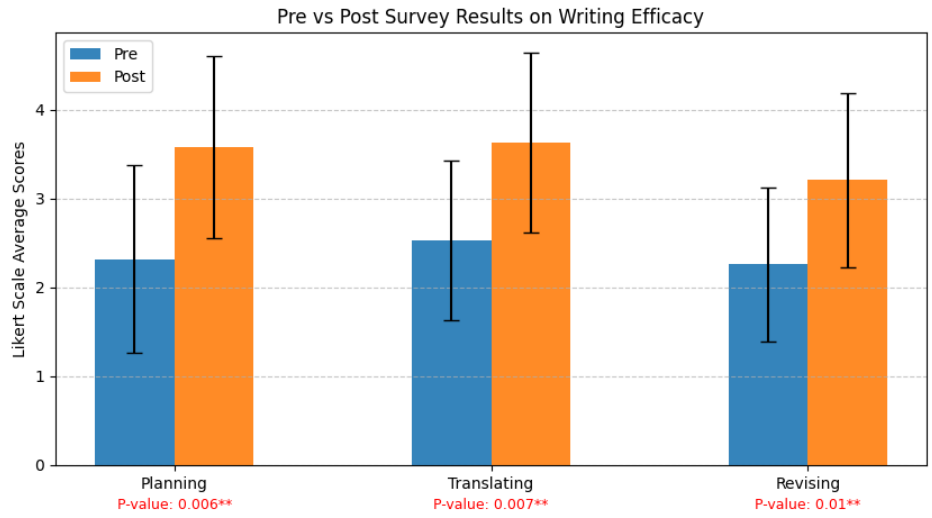


Figure 4.5: Pre-survey results on the English language proficiency, writing habit.

The results of gains in their perception of writing efficacy from the pretest to post-tests demonstrated an improvement for students in all three key competencies (i.e., Planning, Translating, and Revising). Among the three competencies, we found StoryAI improved students' planning stage (i.e., ideation, structure) the most among the other two.

**RQ. Would writing with StoryAI help students develop AI literacy?**

The pre-and-post survey results demonstrated that students' AI literacy has developed, in terms of three aspects of AI literacy (i.e., awareness of what AI is, confidence, and motivation

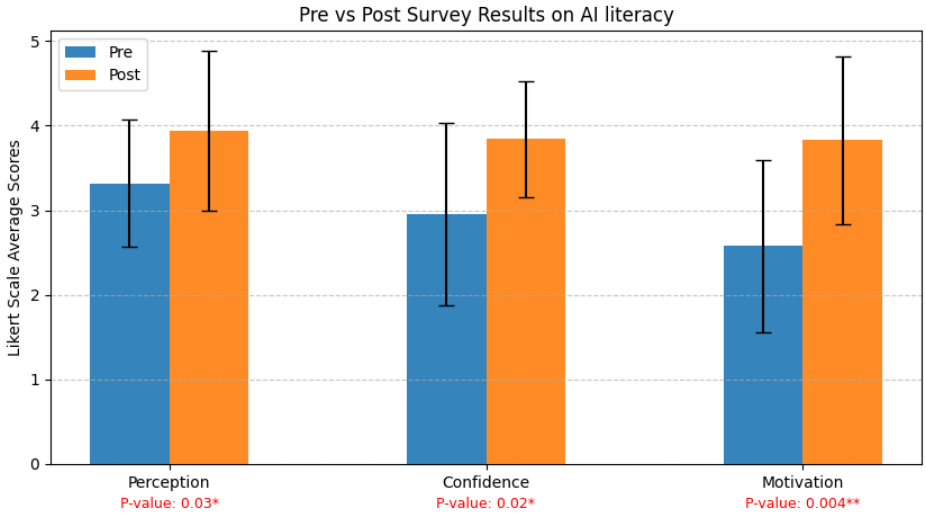


Figure 4.6: Pre-survey results on AI knowledge.

to use AI for their lives) [126]. First, their perceptions about AI (i.e., I think AI is productive and helpful), students' confidence in using AI (i.e., I am confident in using AI), and their motivation (i.e., I want to know and use AI more) were evaluated. All three areas of AI literacy have seen significant improvements. This is especially positive as it enhances awareness and increases young people's experience with AI applications [114]. This data demonstrates that StoryAI not only works to enhance writing efficacy but also raises awareness about AI and broadens participation in understanding AI applications.

### 4.8.2 Usability Evaluation Survey (Post-test)

After students finished writing with StoryAI for two days, students were asked to answer the usability evaluation questionnaires (see Figure 4.7). The survey is designed to understand students' sense of ownership towards their writing output, their StoryAI experience such as enjoyment, and ease of use, and students' opinions about their perception of AI's role as a collaborator, as well as their satisfaction over their writing output in the Likert scale survey. Data demonstrated that the majority of the students (18/30) were satisfied with their writing output, half of the students (15/15) felt they were collaborating with AI agents in StoryAI,

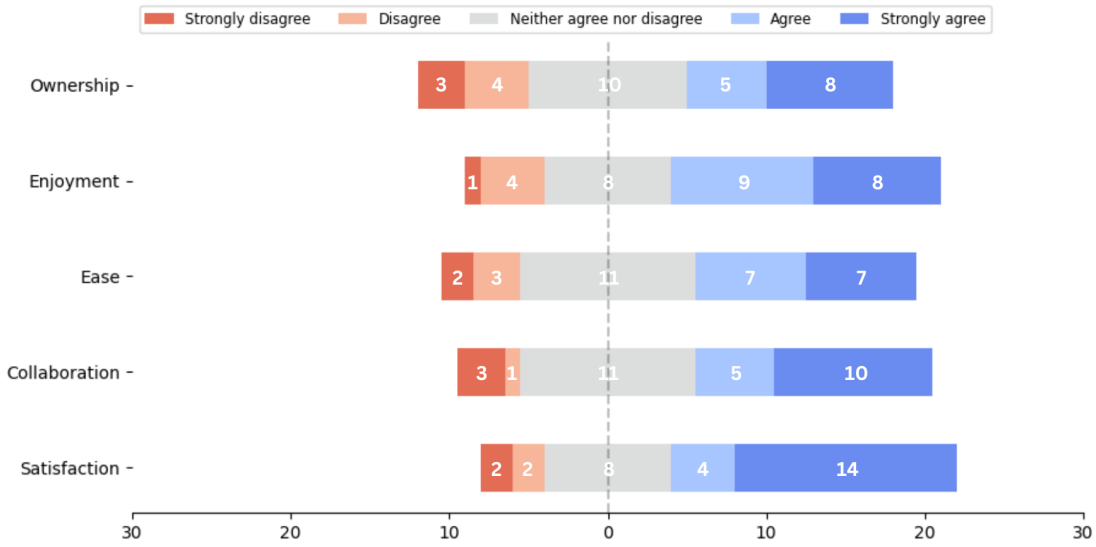


Figure 4.7: Usability Evaluation Survey Result.

fourteen students (14/30) felt that StoryAI was easy to use, although the answers vary due to the age differences in the participants as some of the students who were less literate had a hard time navigating the application. Most students (17/30) enjoyed writing using StoryAI, also shortly less than half (13/30) of the students felt they had a sense of ownership over their writing output. Due to the limitations of the Likert Scale survey, students do not provide detailed explanations for their responses. Therefore, we will offer a more in-depth analysis of these responses in the qualitative analysis section (Section 4.8.3).

### 4.8.3 Qualitative result

**RQ. How do students perceive StoryAI AI agents through their writing?**

Figure 4.8 shows our three high-level categories: affordance of StoryAI for writing education, efficacy, and student perception. The three high-level categories have three sub-categories underneath each theme, motivation, engagement, learning, planning, translating, revising,

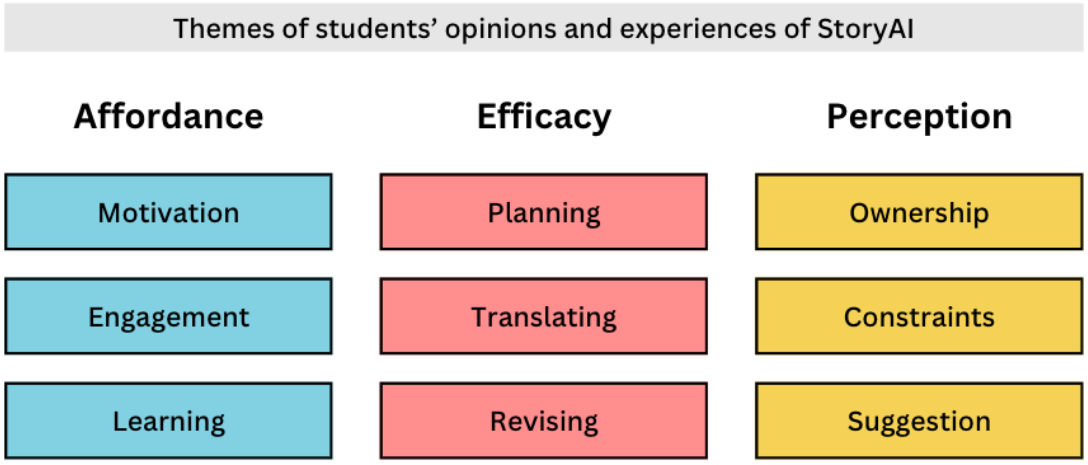


Figure 4.8: Themes about Students' Opinions on StoryAI experiences.

ownership of students' writing output, constraints of StoryAI experiences, and some suggestions on system improvement (see Figure 4.8).

### Affordance of StoryAI in writing education

We describe StoryAI's benefits in terms of how it supports writing activities. Initially, we classify its usefulness in motivation, by grouping students' feedback into sub-categories such as ease of use, enjoyment, and helpfulness. For example, S10 stated StoryAI helps students to initiate their writing easily,

*"StoryAI helps you take the first steps."*

S2 mentioned that StoryAI guides students to plan out a story and it makes them feel easy and fun.

*"I liked that it can succinctly create a beginning and end to the story, which makes it more fun to fill in the center."*

Features such as the ability to add multimedia, customize the story, and provide feedback, were seen as engaging by students. S30 stated,

*"My favorite feature was when the bot gave multiple suggestions to choose from because it let me use more of my own creativity when I needed help with ideas."*

In terms of engagement, we discovered students experienced a personal connection with AI agents as writing partners, communicating their ideas collaboratively. Interview data revealed that the majority of the students felt StoryAI collaborated with them on writing (18/30) and they were happy about their writing output with StoryAI (27/30). We categorized sub-themes with compliment, engaging, attention, and partner. Students mentioned that they think an AI agent is actually paying attention to them as follows,

*"AI engaged me with compliments and showed me like the AI is paying attention to what I'm saying."*

In terms of what they learned from StoryA (learning affordance), students mentioned genre-specific writing, writing structure, grammar, and spelling, specifically, S21 mentioned, *"It tells you like you can check my punctuation and my spelling and grammar errors."*

Finally, when it comes to learning, StoryAI was seen as a tool to help students better understand the writing process and provide a platform for them to practice writing.

**Efficacy of StoryAI in writing project**

We identified a notable benefit of StoryAI for writing in the process of planning, translating, and reviewing students' writing. Among these three phases, StoryAI is most effective during the planning phase of writing. We analyzed student feedback, which focused on 'ideation, planning, and structure,' and found that its benefits are most pronounced at this stage. For example, S12 mentioned,

*"It helped me develop the story step by step. I really do like how it explains to you and it gives you ideas and it makes you think. And then once you're one, you're like, okay, let's*

*move on to the next thing."*

S23 noted that StoryAI helps them craft their stories, adding detail to achieve a better quality of their work,

*"I liked how it was able to create so many ideas because of all of the questions that it asked. It gave me ideas about small details that I probably would not have thought of before."*

As for translation, which involves students translating ideas into sentences, students reported that StoryAI helped add detail, describe, and narrate their ideas. As S20 mentioned,

*"I liked that it occasionally selected and used some specific details that I added. I mentioned daisies in one sentence and they became "vibrant colors of the daisies" in the following AI sentence."*

S9 stated that StoryAI helps them improve the quality of their writing,

*"I think I would use it to become more familiar with the idea of 'show, not tell' when story writing. The AI uses a lot of description and imagery that I would love to use more of."*

StoryAI was also supportive in the revising stage, where students found feedback features useful to quickly and easily refine their writing. Additionally, the AI provided students with detailed feedback on their writing, which allowed them to better understand their strengths and weaknesses (see Figure 4.8).

**Perception towards StoryAI experience**

Students' perception of their StoryAI experience for their writing was very valuable in uncovering some hidden aspects of students' perception. Such as students' sense of ownership of their writing output, we asked questions about who takes authorship of their writing output. Even though all of the students were satisfied with their writing output (27/30), some

of the students mentioned (8/30) they took some AI's ideas (as AI suggestions), therefore, they will give some partial credit to AI. Despite StoryAI having helped the students improve their writing and making them feel as though they owned their work, the students were still willing to acknowledge AI's contribution to their writing. S15 declared,

*"I think the writing output is AI's because I just put the descriptions and it gives me pictures, ideas, and all that, I took ideas from AI."*

Another student S18 mentioned,

*"I feel like, it's both because you could pick the genre and how you want it to look like and AI just does, like gives you ideas and how to finish stuff like that."*

In addition, students also discussed some of the constraints of using StoryAI for writing education, one of the students mentioned she would be willing to pick Google Docs over StoryAI for writing, as she wants to be the main driver of story creation, S7 commented that StoryAl does not provide the same level of customization that Google Docs does. She also mentioned that StoryAl is not as intuitive as Google Docs when it comes to writing. Another student mentioned that he would prefer to use StoryAI as a source of inspiration rather than as a tool for writing, as he prefers to have more control over the creative process. He also noted that AI can be a useful tool to help with writer's block, but he prefers to write his own story from scratch.

*"So then I have my freedom to do my own writing."*

Also, S17 pointed out the potential possibility that people might rely on AI so that when they're no longer available they anxious by saying,

*"You don't have to rely on it too much because you can't use AI whenever you take a test."*

Students also offered some design suggestions of areas to improve StoryAI, one of the students

mentioned color-coded AI-generated output different from students' own writing to alert the AI reliance,

*"I think a feature that can be added is guided colors to see which parts are written by AI and which parts are written by the user. This could help distinguish the story better, as I wanted to clearly see what I wrote."*

Students suggested improvements not only to the interface but also to the instructional design to better support independent learning experiences and enhance AI-mediated conversations. Specifically, they recommended creating opportunities for students to reflect on their AI-assisted work as crucial. This includes encouraging students to analyze their mistakes and consider how they might improve their approach to future tasks through AI-assisted guidance. This includes providing opportunities for student-led learning by enabling students to use AI-assisted tools to create their own independent learning experiences. As S01 said,

*"I think it would be better to help develop students' writing skills if it made prompt sentences, where it starts the first half and then the student fills in from there."*

This would help to challenge students to think critically and to come up with their own ideas. It would also give them the opportunity to explore topics more deeply and to develop their own writing.

In terms of features and functions, many of the students (12/30) recommended a better image-generation feature for their stories. They want AI to be more personalized, scaffolded, and contextually appropriate AI interaction which we will disseminate the insight in the discussion section further.

| Code | Example Quotes |
|---|---|
| **Affordance** | |
| Motivation | "I liked that it can succinctly create a beginning and end to the story, which makes it more fun to fill in the center." |
| Learning | "It helps me learn what genre it is, it matches your genre and how you want it to be." |
| Engagement | "I never knew what a difference a writing partner or bot can make. It's like the rubber ducky technique in coding." |
| **Efficacy** | |
| Planning | "It's good for helping you come up with ideas or to better flesh out your ideas using a sounding board." |
| Translating | "I learned more about, like, narrating my ideas. I like that it was able to guide me along the writing journey." |
| Revising | "It gave me feedback to add more details or if punctuation wasn't right, and the feedback by grading was helpful too." |
| **Perception** | |
| Constraints | "It was ok? It was fun for a minute but it felt like the AI was pushing me towards a predestined route?" |
| Ownership | "Is it mainly my idea? Like I sort of like the details and what you thought about what, like the detail of the detail, you add the detail," |
| Suggestion | "I think a feature that can be added is guided colors to see which parts are written by AI and which parts are written by the user. This could help distinguish the story better, as I wanted to clearly see what I wrote." |

Table 4.3: Code Description and Example Quotes.

## 4.9 Discussion

The study examined the process of designing, developing, and evaluating StoryAI, a generative AI-powered story-authoring platform for children. Despite the system's advantages (i.e., improving students' motivation and efficiency in writing), there are still areas requiring further improvement, such as students' sense of ownership of their writing and students' independent learning and active engagement. In this section, we discuss three key areas that emerged from the study; 1) insights for system designers/developers, 2) insights for researchers, and 3) insights for educators. First, we will address design implications for the future iteration of the platform. Second, we will outline research areas that need to be further examined in order to obtain a deeper understanding of AI-assisted learning tools for their

efficacy and safety. Lastly, we will offer guidance to educators on effectively incorporating AI-assisted learning tools into existing educational practices and ensuring their use is ethical and responsible.

### 4.9.1 Implications for system designers and developers

The results of the study indicate that certain guardrails are necessary for students to remain motivated and engaged in AI-assisted learning. The current system displays students' AI reliance status by leveraging text-similarity API, alerting students' level of AI-generated text being used into percentages (e.g., "You copied 30% of my text"). Along with this approach, we suggest implementing color-coded techniques to highlight what parts are from AI and what parts are from students' own writing in order to provide students with more clarity and understanding about their work and make it easier for teachers to identify AI reliance on students' work. Additionally, it would also help students become more conscious of their reliance on AI in their work and also help students better understand the types of changes they are making to their work. Aligned with that approach, users will benefit from features that allow them to adjust the level of AI assistance. This would help users gain more control over their work, allowing them to customize how much AI assistance they receive. For example, users should be able to adjust the level of AI assistance in their work, from full automation to manual input. This will enable users to tailor their experience to their needs and preferences.

### 4.9.2 Implications for researchers

As we have observed, there are areas of further research that need to be addressed, first, we found that there are still gray areas in examining the efficacy of AI-assisted learning applications regarding the cognitive aspect of students' learning (impact on the short-term

and long-term memory) as well as the writing abilities (i.e., quality of writing such as lexical diversity and sentence count). These can be further studied through longitudinal studies that conduct one-month or one semester-long intervention of AI-assisted writing applications to measure students' writing quality improvement such as lexical diversity, sentence count, and syntax [121, 76]. And short-term long-term memory impact through conducting pre-and post-test measurements, conducting post-tests right after the intervention as well as a one-week or a few weeks afterward. Lastly, researchers will need to understand the impact of students' creativity (i.e., divergent thinking: flexibility, originality, fluency) on students' writing output, which can be examined through randomized control trials by a comparison experiment study where the control group would write without AI-support but the experiment group would be given AI-suggestions on their ideas over their writing. Then researchers analyze and compare the two groups' writing results with TTCT measurement (fluency, originality, flexibility, and elaboration) [167]. Aside from learning, researchers might want to assess and analyze whether they're active or passive in learning engagement. This can be studied by assessing students' engagement using the ICAP framework [35]. Through the course of studies, reporting its findings in research communities would help build trust and reliability of AI-assisted learning applications for educational purposes.

### 4.9.3 Implications for educators

The results of our study highlight the need to increase young people's awareness and involvement with AI systems. As Long & Magerko [114] initiated as a competency – recognizing AI – an important part of the StoryAI experience is being able to differentiate between technological artifacts that use AI and those that do not. In addition to that, to help students to equip a critical understanding of AI's capabilities (i.e., critically analyze and discuss AI), it is important to educate them about AI's capabilities and limitations Furthermore, students should be taught to think critically about the ethical implications of AI. They should

also be encouraged to ask questions and challenge existing assumptions and beliefs. Finally, educators should also provide students with guidance on how to responsibly use AI. This could be accomplished through providing students with the opportunity to develop their own AI-related stories, with AI-related languages (i.e., algorithms, data, machine learning) as this will help them to better understand the technical elements of AI. Educators could also assign students to explore AI-related topics in greater depth, such as analyzing the impact of AI on society. Additionally, they could also assign students to develop their AI-related projects, such as developing an AI-based chatbot or creating a video game with AI. This includes AI literacy education, which can be accomplished in part of STEM learning activities, or through extensively analyzing potential risks and challenges of leveraging the system for education.

### 4.9.4 Limitation

As our study was conducted in out-of-school settings, in one of the non-profit community centers where that predominantly serve Hispanics/Latinos, 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 students in the study were multilingual, all of the students identify themselves as Latinos. Because our samples lack a diverse cultural background, they may have limited perspectives and opinions. Therefore, more diverse samples should be recruited to gain a comprehensive understanding of the platform's impact.

The study was conducted for four days including two days of active use on StoryAI which makes it hard to validate the long-term effect on students' writing proficiency. As a result, our results could not shed light on students' long-term interaction patterns. To gain a more accurate understanding of the platform's efficacy, a longer study with more days of active use is needed which could be possible to answer through longitudinal studies by saving, tracking,

and analyzing students' writing output over time.

Even though we tried to understand writing skills improvement through StoryAI, there is a lack of evidence on the specific writing skills acquisition. Thus, further research is needed to better understand the specific learning outcomes and benefits of StoryAI including improvement in their lexical diversity in writing, increased sentence count, better word choices, grammar, and syntax. This can be answered through randomized control trials to compare groups' writing outcomes, one with a group with AI-assisted tools like StoryAI and one with writing without AI-assisted tools.

Additionally, the platform has several limitations. First, we need to achieve the right balance between AI-assisted, algorithmic guidance and teacher-crafted, instructor-guided learning experiences when designing AI-assisted learning applications. Over-defining the instructional path can restrict students' open-ended conversations with the AI, limiting the flexibility of their interactions. However, it is challenging to mitigate the risk of AI being used in inappropriate or out-of-context situations when lacking the structured educational guidance that follows certain conversation prompts. Furthermore, there are also ethical considerations to consider, such as the potential misuse of AI-assisted tools by students, and the potential bias that AI-assisted tools may bring into the classroom.

Aside from this, the AI-assisted learning application usage is heavily dependent on students' digital literacy, it is important to ensure that students have sufficient digital literacy skills to be able to effectively use the AI-assisted tool. It is also important to ensure that AI-assisted tools are used responsibly and that students are aware of the potential risks associated with their misuse. As part of their responsibilities, educators should provide adequate support for students when it comes to their digital literacy skills, ensure that artificial intelligence-assisted tools are used ethically and responsibly, and take appropriate measures to protect students' privacy.

## 4.10 Conclusion

From the study, we discussed the design, development, and evaluation of StoryAI, a GenAI-powered platform for lesson authoring and learning. This platform integrates a GenAI-enabled chatbot, a mechanism for processing user-defined conversational protocols, and a flexible canvas element. It aims to provide students with an engaging learning experience that mirrors the interactive dynamics of a traditional learning environment. Simultaneously, it offers educators the ability to create adaptable and progressive instructional sequences that the AI agent can carry out. With StoryAI as the foundation for narrative writing activity AI instruction, our goal is to develop a digital platform capable of delivering interactive online lessons accessible to all learners, with a particular emphasis on children, across any subject or body of knowledge. Distinct from conventional online educational resources, our goal is to compile a collection of a variety of educational materials—such as images, manuscripts, textual inquiries, and videos. Instead, we aim to orchestrate a cohesive educational journey that mirrors the dynamics of learning experiences. In this envisioned environment, learners engage with content sequentially, adhering to the instructional guidance provided by educators, thereby fostering an iterative process of knowledge acquisition, reflection, and the development of individual insights.

# Chapter 5

# Conclusion

The dissertation consists of three studies that are the process of designing, developing, and evaluating generative-AI-powered story-authoring platforms for children. The first study focuses on the formative study of how stakeholders in education (i.e., teachers, parents, and students) perceive and leverage generative AI platforms (i.e., ChatGPT and Stable Diffusion) for writing activities. I found that the GenAI systems could be beneficial in generating adaptable teaching materials for teachers, enhancing ideation, and providing students with personalized, timely feedback.

The second study focuses on co-designing a logic model in informing designing AI-based Writing Tutoring Platforms (AWTP) with educators, then designing and evaluating AWTP prototypes that focus on opinion writing. From the co-design process, we identified the platform's potential users, features, functionalities, and desired outcomes. With this insight, we created a prototype, AWTP. The usability study findings with the AWTP prototype suggested AWTP's efficacy in improving students' writing engagement by increasing their time spent in writing, total word count, and lexical diversity. Feedback study revealed AWTP's potential efficacy in improving motivation in writing by reducing anxiety over writing for

emergent writers.

The third study focuses on designing, developing, and evaluating story-authoring platforms, StoryAI, for narrative writing for children. From the usability study, I found StoryAI's efficacy in students' perception of writing competencies (i.e., planning, translating, and revising) as well as AI literacy (i.e., perception, confidence, and motivation).

StoryAI is an ongoing project, with iterative design and development processes to continuously improve its functionality and accessibility. The contribution of the study is to inform the HCI and learning science community by highlighting the practical applications and limitations of GenAI in story writing for children, and by offering insights that can guide the design and implementation of GenAI tools in a way that aligns with the needs and concerns of various educational stakeholders. And, the research will advance the understanding of child-centered AI, emphasizing the importance of developing technologies that are suitable for young learners' cognitive and emotional needs. The research will offer empirical evidence on the applicability of GenAI in educational settings by highlighting new directions for research in GenAI and education, paving the way for further innovations and studies in this rapidly evolving field.

To mitigate the challenges of using GenAI systems in writing projects, such as to mitigate inappropriate content generation or interaction with AI agents, we design and develop the AI agent with teachers, and continue to develop teacher-in-the-loop systems where adults can oversee the child-AI interaction. We also facilitate a database where we can scrutinize student and AI chat logs for further studies around student-AI interactions and learning experiences to inform the improvement of the system's development.

All of these insights have given us recommendations for system developers, researchers, and educators. First, the platform needs teachers' views so that they can adjust and monitor students' AI assistants and manage their instruction. s expressed interest in customizing

the system's difficulty levels to accommodate students with varying abilities. Also, alert systems will be helpful for students to be in the driver's seat in their learning processes, like color-coded AI-generated outputs different from what students write themselves.

Further research is needed in the following areas, as I have observed. First, I found that there are still gray areas in examining the efficacy of AI-assisted learning applications on students' learning (i.e., quality of writing such as lexical diversity and sentence count), creativity, and impact on short-term and long-term memory. To understand the impact of student's creativity (i.e., divergent thinking: flexibility, originality, fluency). This research can be conducted by a comparison experiment study where the control group would write without AI support but the experiment group would be given AI suggestions on their ideas over their writing. Then researchers analyze and compare the two groups' writing results with TTCT measurement, lexical diversity. Also, conduct a post-test shortly after and a week later.

The study highlights the need to increase young people's awareness and involvement with AI systems. It's important to educate them about AI's capabilities and limitations, helping them understand what AI can and cannot do. This includes AI literacy education, which can be accomplished as part of STEM learning activities, As we noticed from the StoryAI study, interacting with AI agents helps students to engage in story creation, it is another way to Integrate AI literacy with storytelling activity. Ultimately, it provides students with guidance on how to use AI-assisted tools effectively.

The research clearly indicates that integrating AI into education requires a dual focus: not only should we utilize AI to enhance educational outcomes, but we must also ensure that users are educated on its safe and effective use. To achieve this, simultaneous training in AI, data literacy, and AI ethics is essential. This comprehensive approach will cultivate trust and guarantee the safe and effective application of AI technology across all educational settings.

The contributions of this research are as follows:

1. The study will provide empirical evidence on the effectiveness of Generative AI (GenAI) in educational settings, thereby enriching our understanding of its advantages and constraints.

2. It will enhance our comprehension of child-centered AI, underscoring the necessity of creating technologies that cater to the cognitive and emotional requirements of young learners.

3. The research will identify novel avenues for investigation in the intersection of GenAI and education, setting the stage for future innovations and informing subsequent integrations of GenAI in educational contexts.

I will wrap the dissertation with the future research direction, as I'm still working on iterating the StoryAI platform, and keep studying its efficacy, a short-term plan would be to optimize the platform via educational behavior data and user testing. I aim to use the platform's educational behavior data to assess and measure students' learning processes and outcomes. This includes model fine-tuning, to enhance its functionality and safety, especially for students with diverse needs. I also would like to broaden subjects to integrate diverse subjects, I strongly believe storytelling is a powerful tool for acquiring new knowledge, so I would like to use the platform's story-creation methods to teach broader subjects, especially STEAM areas, like science learning through storytelling. I'd like to collaborate with educators to design curriculum and instructions to ensure the platform's relevance and effectiveness in their settings. And facilitate teachers' views so educators can optimize and customize lesson plans and moderate AI integration in their classes.

The ultimate goal is to iteratively refine and improve the platform to foster adaptive learning at scale and support child-centered, inclusive educational experiences. Current regulations for self-driving cars require drivers to remain in the driver's seat, take the lead, and stay actively engaged in driving. If the driver lets go of the steering wheel, the car will make

a loud warning beep. This concept is also relevant to AI-assisted educational applications, where students must remain actively engaged in their learning process. While they can utilize automatic features to streamline certain tasks and perhaps find more efficient learning paths, this doesn't mean they should automate their core objective of learning.

Learning is a complex cognitive process that AI cannot fully replicate, particularly when it comes to translating cognitive aspects of student learning. Nevertheless, AI should empower students to take control of their educational journey and encourage critical thinking, rather than automating the learning process. Research should continue to find optimal ways to support students' active learning through AI-assisted learning.

# Bibliography


[1] Login — Bedtimestory.ai — bedtimestory.ai. https://www.bedtimestory.ai/create. [Accessed 13-06-2024].

[2] Naria - Stories Crafted by You — naria.ai. https://naria.ai/. [Accessed 13-06-2024].

[3] Once Upon a Bot &x2022; Create Children's Stories with AI — onceuponabot.com. https://onceuponabot.com/. [Accessed 13-06-2024].

[4] Scarlett Panda — Customized Short Stories App for Children — scarlettpanda.com. https://www.scarlettpanda.com/. [Accessed 13-06-2024].

[5] Story.com — Storytelling Meets AI — story.com. https://www.story.com/. [Accessed 13-06-2024].

[6] StoryNest.ai - AI powered Interactive Novels and Stories — storynest.ai. https://storynest.ai/home. [Accessed 13-06-2024].

[7] Storywizard.ai — Create incredible learning experiences using AI — storywizard.ai. https://www.storywizard.ai/. [Accessed 13-06-2024].

[8] wemadeastory.com — Generate original children's stories with the help of Artificial Intelligence — wemadeastory.com. https://www.wemadeastory.com/login. [Accessed 13-06-2024].

[9] F. Agostinelli, M. Mavalankar, V. Khandelwal, H. Tang, D. Wu, B. Berry, B. Srivastava, A. Sheth, and M. Irvin. 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*, IDC '21, page 610–614, New York, NY, USA, 2021. Association for Computing Machinery.

[10] N. Ahmad, S. Murugesan, and N. Kshetri. Generative Artificial Intelligence and the Education Sector. *Computer*, 56(6):72–76, June 2023. Conference Name: Computer.

[11] M. ai. Ai for teachers - lesson planning and more!, 2023.

[12] O. AI. OpenAI API, 2023.

[13] S. ai. The speaking app that actually talks, 2023.

[14] M. Alaimi, E. Law, K. D. Pantasdo, P.-Y. Oudeyer, and H. Sauzeon. Pedagogical agents for fostering question-asking skills in children. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[15] C. Angevine, K. Cator, B. Liberman, K. Smith, and V. Young. Designing a Process for Inclusive Innovation: A Radical Commitment to Equity. Technical report, Digital Promise, Nov. 2019.

[16] V. application. Build amazing conversational assistants, 2023.

[17] B. Aronson and J. Laughter. The Theory and Practice of Culturally Relevant Education: A Synthesis of Research Across Content Areas. *Review of Educational Research*, 86(1):163–206, Mar. 2016.

[18] D. Baidoo-Anu and L. Owusu Ansah. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning, Jan. 2023.

[19] L. B. Baruque and R. N. Melo. Learning theory and instruction design using learning objects. *Journal of Educational Multimedia and Hypermedia*, 13(4):343–370, 2004.

[20] R. Beard, J. Riley, and D. Myhill. The sage handbook of writing development. 2009.

[21] E. Beneteau, O. K. Richards, M. Zhang, J. A. Kientz, J. Yip, and A. Hiniker. Communication breakdowns between families and alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery.

[22] O. C. Biermann, N. F. Ma, and D. Yoon. From Tool to Companion: Storywriters Want AI Writers to Respect Their Personal Values and Writing Strategies. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, pages 1209–1227, New York, NY, USA, June 2022. Association for Computing Machinery.

[23] G. Blockly, 2023), publisher=Google.

[24] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

[25] O. U. a. Bot. Once upon a bot, 2023.

[26] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language Models are Few-Shot Learners, July 2020. arXiv:2005.14165 [cs].

[27] P. Brusilovsky. Ai in education, learner control, and human-ai collaboration. *International Journal of Artificial Intelligence in Education*, 34(1):122–135, 2024.

[28] B. Cagiltay, B. Mutlu, and J. E. Michaelis. "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*, IDC '23, page 375–387, New York, NY, USA, 2023. Association for Computing Machinery.

[29] R. Cappelli. *Mentor texts: Teaching writing through children's literature, K-6*. Routledge, 2023.

[30] M. A. Cardona, R. J. Rodríguez, and K. Ishmael. Artificial Intelligence and the Future of Teaching and Learning. 2023.

[31] N. Carlini, F. Tramer, E. Wallace, M. Jagielski, A. Herbert-Voss, K. Lee, A. Roberts, T. Brown, D. Song, U. Erlingsson, et al. Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 2633–2650, 2021.

[32] ChatGPT. Chatgpt, 2023.

[33] ChatGPT. Introducing ChatGPT, 2023.

[34] L. Chen, P. Chen, and Z. Lin. Artificial Intelligence in Education: A Review. *IEEE Access*, 8:75264–75278, 2020. Conference Name: IEEE Access.

[35] M. T. Chi and R. Wylie. The icap framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, 49(4):219–243, 2014.

[36] M. T. H. Chi and R. Wylie. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4):219–243, Oct. 2014. Publisher: Routledge _eprint: https://doi.org/10.1080/00461520.2014.965823.

[37] T. K. F. Chiu, Q. Xia, X. Zhou, C. S. Chai, and M. Cheng. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4:100118, Jan. 2023.

[38] I.-A. Chounta, E. Bardone, A. Raudsep, and M. Pedaste. 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):725–755, 2022.

[39] E. Clark, A. S. Ross, C. Tan, Y. Ji, and N. A. Smith. Creative writing with a machine in the loop: Case studies on slogans and stories. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, page 329–340, New York, NY, USA, 2018. Association for Computing Machinery.

[40] D. R. Cotton, P. A. Cotton, and J. R. Shipway. Chatting and cheating: Ensuring academic integrity in the era of chatgpt. *Innovations in Education and Teaching International*, pages 1–12, 2023.

[41] CreateBookAI. Create Book. `https://createbookai.com/`. [Accessed 13-06-2024].

[42] M. Crowhurst. Teaching and learning the writing of persuasive/argumentative discourse. *Canadian Journal of Education/Revue canadienne de l'éducation*, pages 348–359, 1990.

[43] M. Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. Harper Collins, Oct. 2009. Google-Books-ID: QVjPsd1UukEC.

[44] L. Cuban. *Oversold and underused: Computers in the classroom*. Harvard university press, 2001.

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

[46] L. Cutler and S. Graham. Primary grade writing instruction: A national survey. *Journal of educational Psychology*, 100(4):907, 2008.

[47] C. M. Czerniak and J. J. Haney. The effect of collaborative concept mapping on elementary preservice teachers' anxiety, efficacy, and achievement in physical science. *Journal of Science Teacher Education*, 9(4):303–320, 1998.

[48] W. Damon. Peer education: The untapped potential. *Journal of applied developmental psychology*, 5(4):331–343, 1984.

[49] N. Dehouche. Plagiarism in the age of massive generative pre-trained transformers (gpt-3). *Ethics in Science and Environmental Politics*, 21:17–23, 2021.

[50] P. DelSignore. Train your own image model using leonardo ai, Apr 2023.

[51] G. Dietz, J. K. Le, N. Tamer, J. Han, H. Gweon, E. L. Murnane, and J. A. Landay. StoryCoder: Teaching Computational Thinking Concepts Through Storytelling in a Voice-Guided App for Children. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–15, New York, NY, USA, 2021. Association for Computing Machinery.

[52] S. Diffusion. Stable diffusion online, 2023.

[53] S. Diffusion. Stable Diffusion Public Release, 2023.

[54] P. Digital. Learner variability navigator: Learner variability project, 2021.

[55] C. Drahl. Ai was asked to create images of black african docs treating white kids. how'd it go?, Oct 2023.

[56] S. Druga, F. L. Christoph, and A. J. Ko. 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*, pages 1–17, 2022.

[57] J. Du and B. K. Daniel. Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, 6:100230, June 2024.

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

[59] M. Elgarf, S. Zojaji, G. Skantze, and C. Peters. CreativeBot: a Creative Storyteller robot to stimulate creativity in children. In *Proceedings of the 2022 International Conference on Multimodal Interaction*, ICMI '22, pages 540–548, New York, NY, USA, 2022. Association for Computing Machinery.

[60] J. E. Fischer. Generative ai considered harmful. In *Proceedings of the 5th International Conference on Conversational User Interfaces*, CUI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[61] V. Flow. vueflow: the customizable vue3 flowchart library, 2023.

[62] L. Flower and J. R. Hayes. A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4):365–387, 1981. Publisher: National Council of Teachers of English.

[63] G. Franceschelli and M. Musolesi. On the creativity of large language models. *arXiv preprint arXiv:2304.00008*, 2023.

[64] R. Garg, H. Cui, S. Seligson, B. Zhang, M. Porcheron, L. Clark, B. R. Cowan, and E. Beneteau. 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*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[65] R. Garg and S. Sengupta. 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*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[66] R. Garg and S. Sengupta. 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*, pages 1–13, 2020.

[67] K. I. Gero, V. Liu, and L. Chilton. Sparks: Inspiration for science writing using language models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, page 1002–1019, New York, NY, USA, 2022. Association for Computing Machinery.

[68] K. I. Gero, V. Liu, and L. Chilton. Sparks: Inspiration for Science Writing using Language Models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, pages 1002–1019, New York, NY, USA, June 2022. Association for Computing Machinery.

[69] K. I. Gero, T. Long, and L. B. Chilton. Social dynamics of ai support in creative writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[70] J. Gilbert and S. Graham. Teaching writing to elementary students in grades 4–6: A national survey. *The elementary school journal*, 110(4):494–518, 2010.

[71] A. C. Graesser, D. S. McNamara, Z. Cai, M. Conley, H. Li, and J. Pennebaker. Coh-metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal*, 115(2):210–229, 2014.

[72] A. C. Graesser, D. S. McNamara, and J. M. Kulikowich. Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, 40(5):223–234, June 2011. Publisher: American Educational Research Association.

[73] S. Graham, A. Bollinger, C. B. Olson, C. D'Aoust, C. MacArthur, D. McCutchen, and N. Olinghouse. Teaching elementary school students to be effective writers: A practice guide. ncee 2012-4058. *What Works Clearinghouse*, 2012.

[74] S. Graham, K. R. Harris, E. Fishman, J. Houston, K. Wijekumar, P.-W. Lei, and A. B. Ray. Writing skills, knowledge, motivation, and strategic behavior predict students' persuasive writing performance in the context of robust writing instruction. *The Elementary School Journal*, 119(3):487–510, 2019.

[75] S. Graham, K. R. Harris, S. A. Kiuhara, and E. J. Fishman. The relationship among strategic writing behavior, writing motivation, and writing performance with young, developing writers. *The Elementary School Journal*, 118(1):82–104, 2017.

[76] S. Graham, M. Hebert, M. Paige Sandbank, and K. R. Harris. Assessing the writing achievement of young struggling writers: Application of generalizability theory. *Learning Disability Quarterly*, 39(2):72–82, 2016.

[77] S. Graham and D. Perin. A meta-analysis of writing instruction for adolescent students. *Journal of educational psychology*, 99(3):445, 2007.

[78] S. Graham and D. Perin. Writing next-effective strategies to improve writing of adolescents in middle and high schools, 2007.

[79] S. Graham, K. Wijekumar, K. R. Harris, P.-W. Lei, E. Fishman, A. B. Ray, and J. Houston. Writing skills, knowledge, motivation, and strategic behavior predict students' persuasive writing performance in the context of robust writing instruction. *The Elementary School Journal*, 119(3):487–510, Mar 2019.

[80] A. Gupta and C. J. MacLellan. Designing teachable systems for intelligent tutor authoring. 2021.

[81] A. Han and Z. Cai. Design implications of generative ai systems for visual storytelling for young learners. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, IDC '23, page 470–474, New York, NY, USA, 2023. Association for Computing Machinery.

[82] A. Han, X. Zhou, Z. Cai, S. Han, R. Ko, S. Corrigan, and K. Peppler. Teachers, Parents, and Students' Perspectives on Integrating Generative AI into Elementary Literacy Education. 2024.

[83] K. R. Harris and S. Graham. Self-regulated strategy development in writing: Premises, evolution, and the future. *British Journal of Educational Psychology*, 2(6):113–135, 2009.

[84] T. Haşlaman, F. K. Mumcu, and N. A. Uslu. Fostering computational thinking through digital storytelling: a distinctive approach to promoting computational thinking skills of pre-service teachers. *Education and Information Technologies*, pages 1–27, 2024.

[85] P. B. Horton, A. A. McConney, M. Gallo, A. L. Woods, G. J. Senn, and D. Hamelin. An investigation of the effectiveness of concept mapping as an instructional tool. *Science education*, 1993.

[86] M. Huang, X. Zhu, and J. Gao. Challenges in building intelligent open-domain dialog systems. *ACM Trans. Inf. Syst.*, 38(3), apr 2020.

[87] B. Huber, D. McDuff, C. Brockett, M. Galley, and B. Dolan. Emotional Dialogue Generation using Image-Grounded Language Models. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–12, New York, NY, USA, Apr. 2018. Association for Computing Machinery.

[88] S. Hunter, J. Kalanithi, and D. Merrill. 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, pages 206–209, New York, NY, USA, 2010. Association for Computing Machinery.

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

[90] G.-J. Hwang, H. Xie, B. W. Wah, and D. Gašević. Vision, challenges, roles and research issues of artificial intelligence in education, 2020.

[91] IES. Institute of education sciences, 2023.

[92] A. Islam and D. Inkpen. Semantic text similarity using corpus-based word similarity and string similarity. *ACM Trans. Knowl. Discov. Data*, 2(2), jul 2008.

[93] S. L. Jaeho Jeon and S. Choi. A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments*, 0(0):1–19, 2023.

[94] J. Jeon. Exploring AI chatbot affordances in the EFL classroom: young learners' experiences and perspectives. *Computer Assisted Language Learning*, 37(1-2):1–26, Jan. 2024. Publisher: Routledge _eprint: https://doi.org/10.1080/09588221.2021.2021241.

[95] S. Kalantari, E. Rubegni, L. Benton, and A. Vasalou. "when i'm writing a story, i am really good" exploring the use of digital storytelling technology at home. *International Journal of Child-Computer Interaction*, 38:100613, 2023.

[96] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, et al. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274, 2023.

[97] A. Kawalkar and J. Vijapurkar. Scaffolding science talk: The role of teachers' questions in the inquiry classroom. *International Journal of Science Education*, 35(12):2004–2027, 2013.

[98] R. T. Kellogg and B. A. Raulerson. Improving the writing skills of college students. *Psychonomic bulletin & review*, 14:237–242, 2007.

[99] M. Khalil and E. Er. Will chatgpt get you caught? rethinking of plagiarism detection. *arXiv preprint arXiv:2302.04335*, 2023.

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

[101] A. Klein. 180 Degree Turn: NYC District Goes From Banning ChatGPT to Exploring AI's Potential. *Education Week*, Oct. 2023.

[102] D. Kristen. Khan academy explores the potential for gpt-4 in a limited pilot program., 2023.

[103] N. Kucirkova. Children's agency by design: Design parameters for personalization in story-making apps. *International Journal of Child-Computer Interaction*, 21:112–120, 2019.

[104] M. Lee, P. Liang, and Q. Yang. 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, pages 1–19, New York, NY, USA, Apr. 2022. Association for Computing Machinery.

[105] W. M. Lim, A. Gunasekara, J. L. Pallant, J. I. Pallant, and E. Pechenkina. Generative ai and the future of education: Ragnarök or reformation? a paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2):100790, 2023.

[106] W. M. Lim, A. Gunasekara, J. L. Pallant, J. I. Pallant, and E. Pechenkina. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2):100790, July 2023.

[107] P. Lin and J. Van Brummelen. 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, pages 1–12, New York, NY, USA, May 2021. Association for Computing Machinery.

[108] P. Lin, J. Van Brummelen, G. Lukin, R. Williams, and C. Breazeal. Zhorai: Designing a conversational agent for children to explore machine learning concepts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13381–13388, 2020.

[109] M. Lindh and J. Nolin. Information we collect: Surveillance and privacy in the implementation of google apps for education. *European Educational Research Journal*, 15(6):644–663, 2016.

[110] C.-C. Liu, M.-G. Liao, C.-H. Chang, and H.-M. Lin. An analysis of children'interaction with an ai chatbot and its impact on their interest in reading. *Computers & Education*, 189:104576, 2022.

[111] V. Liu and L. B. Chilton. 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, pages 1–23, New York, NY, USA, Apr. 2022. Association for Computing Machinery.

[112] V. Liu, H. Qiao, and L. Chilton. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, pages 1–17, New York, NY, USA, Oct. 2022. Association for Computing Machinery.

[113] D. Long, T. Blunt, and B. Magerko. Co-designing ai literacy exhibits for informal learning spaces. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–35, 2021.

[114] D. Long and B. Magerko. What is ai literacy? competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–16, New York, NY, USA, 2020. Association for Computing Machinery.

[115] B. Loveless. Learning to Read to Read to Learn. Myth or Reality?, 2023.

[116] J. Low. A pragmatic definition of the concept of theoretical saturation. *Sociological Focus*, 52(2):131–139, Jan 2019.

[117] W. ly. Wanderly: Magical adventures for curious children, 2023.

[118] C. J. MacLellan, R. Sheline, and E. Harpstead. A human-centered approach to designing teachable systems. 2019.

[119] D. Malvern, B. Richards, N. Chipere, and P. Durán. *Lexical diversity and language development*. Springer, 2004.

[120] P. M. McCarthy and S. Jarvis. MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2):381–392, May 2010.

[121] D. S. McNamara, S. A. Crossley, and P. M. McCarthy. Linguistic Features of Writing Quality. *Written Communication*, 27(1):57–86, Jan. 2010. Publisher: SAGE Publications Inc.

[122] Midjourney. Midjourney, 2023.

[123] S. Miller and L. Pennycuff. The power of story: Using storytelling to improve literacy learning. *Journal of Cross-Disciplinary Perspectives in Education*, 1(1):36–43, 2008.

[124] M. Muller, L. B. Chilton, A. Kantosalo, C. P. Martin, and G. Walsh. GenAICHI: Generative AI and HCI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, pages 1–7, New York, NY, USA, 2022. Association for Computing Machinery.

[125] J. C. Nesbit and O. O. Adesope. Learning with concept and knowledge maps: A meta-analysis. *Review of educational research*, 76(3):413–448, 2006.

[126] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. S. Qiao. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2:100041, Jan. 2021.

[127] H. Nguyen. Role design considerations of conversational agents to facilitate discussion and systems thinking. *Computers & Education*, 192:104661, 2023.

[128] H. Nguyen. Role design considerations of conversational agents to facilitate discussion and systems thinking. *Computers & Education*, 192:104661, Jan. 2023.

[129] E. Nichols, L. Gao, and R. Gomez. Collaborative storytelling with large-scale neural language models. In *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pages 1–10, 2020.

[130] J. Nordlof. Vygotsky, scaffolding, and the role of theory in writing center work. *The Writing Center Journal*, pages 45–64, 2014.

[131] OpenAI. DALL·E 2, 2022.

[132] OpenAI. Introducing gpts, 2023.

[133] D. Paris and H. S. Alim. *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world*. Teachers College Press, 2017.

[134] P. Pataranutaporn, V. Danry, L. Blanchard, L. Thakral, N. Ohsugi, P. Maes, and M. Sra. Living memories: Ai-generated characters as digital mementos. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23, page 889–901, New York, NY, USA, 2023. Association for Computing Machinery.

[135] A. Peng, M. Wu, J. Allard, L. Kilpatrick, and S. Heidel, 2023.

[136] K. Peppler. New opportunities for interest-driven arts learning in a digital age. *Wallace Foundation*, 2013.

[137] K. Peppler, R. Santo, M. Gresalfi, and K. S. Tekinbas. *Script changers: digital storytelling with Scratch*. MIT Press, 2014.

[138] I. Pesek, N. Nosović, and M. Krašna. The role of ai in the education and for the education. In *2022 11th Mediterranean Conference on Embedded Computing (MECO)*, pages 1–4. IEEE, 2022.

[139] J. Piaget. *Piaget's theory of intelligence*. Englewood Cliffs, NJ: Prentice Hall, 1978.

[140] S. Polak, G. Schiavo, and M. Zancanaro. Teachers' perspective on artificial intelligence education: An initial investigation. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA, 2022. Association for Computing Machinery.

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

[142] C. Y. Quah and K. H. Ng. A systematic literature review on digital storytelling authoring tool in education: January 2010 to january 2020. *International Journal of Human–Computer Interaction*, 38(9):851–867, 2022.

[143] C. T. Rafal. From co-construction to takeovers: Science talk in a group of four girls. *The Journal of the Learning Sciences*, 5(3):279–293, 1996.

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

[145] L. J. Rennie and T. Jarvis. Three approaches to measuring children's perceptions about technology. *International Journal of Science Education*, 17(6):755–774, 1995.

[146] L. J. Rennie and T. Jarvis. Three approaches to measuring children's perceptions about technology. *International Journal of Science Education*, 17(6):755–774, 1995.

[147] M. Resnick, B. Myers, K. Nakakoji, B. Shneiderman, R. Pausch, and M. Eisenberg. Design Principles for Tools to Support Creative Thinking. *Report of Workshop on Creativity Support Tools*, 20, Jan. 2005.

[148] K. Rosenblatt. Chatgpt banned from new york city public schools' devices and networks, Jan 2023.

[149] K. Rosenblatt. New york city public schools remove chatgpt ban, May 2023.

[150] D. Russell, Q. V. Liao, C. Kulkarni, E. L. Glassman, and N. Martelaro. 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*, CHI EA '23, New York, NY, USA, 2023. Association for Computing Machinery.

[151] J. Saldaña. *The coding manual for qualitative researchers*. SAGE, Los Angeles, 2nd ed edition, 2013. OCLC: ocn796279115.

[152] I. T. Sanusi, S. S. Oyelere, H. Vartiainen, J. Suhonen, and M. Tukiainen. A systematic review of teaching and learning machine learning in k-12 education. *Education and Information Technologies*, 28(5):5967–5997, 2023.

[153] O. Schmitt and D. Buschek. CharacterChat: Supporting the Creation of Fictional Characters through Conversation and Progressive Manifestation with a Chatbot. In *Proceedings of the 13th Conference on Creativity and Cognition*, C&C '21, pages 1–10, New York, NY, USA, June 2021. Association for Computing Machinery.

[154] K. Sherwin. Card sorting: Uncover users' mental models for better information architecture, Mar 2018.

[155] R. Shrivastava. Amazon alexa wants to put your child to bed with generative ai storytelling, Dec 2022.

[156] N. Singh, G. Bernal, D. Savchenko, and E. L. Glassman. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Trans. Comput.-Hum. Interact.*, feb 2022. Just Accepted.

[157] M. Skjuve, A. Følstad, K. I. Fostervold, and P. B. Brandtzaeg. My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 149:102601, 2021.

[158] P. Smutny and P. Schreiberova. Chatbots for learning: A review of educational chatbots for the facebook messenger. *Computers & Education*, 151:103862, 2020.

[159] K. C. Stegemann and A. P. Jaciw. Making it logical: Implementation of inclusive education using a logic model framework. *Learning Disabilities: A Contemporary Journal*, 16(1):3–18, 2018.

[160] J. Steiss, T. Tate, S. Graham, J. Cruz, M. Hebert, J. Wang, Y. Moon, W. Tseng, et al. Comparing the quality of human and chatgpt feedback on students' writing. 2023.

[161] J. Su, D. T. K. Ng, and S. K. W. Chu. Artificial intelligence (ai) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence*, 4:100124, 2023.

[162] B. Swanson, K. Mathewson, B. Pietrzak, S. Chen, and M. Dinalescu. 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*, pages 244–256, 2021.

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

[164] I. Text. Analyse the difficulty level of english texts, Apr 2023.

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

[166] A. ti. The #1 software for qualitative data analysis, Sep 2023.

[167] E. P. Torrance. Torrance tests of creative thinking. *Educational and psychological measurement*, 1966.

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

[169] J. Van Brummelen, M. Kelleher, M. C. Tian, and N. Nguyen. 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*, IDC '23, page 187–197, New York, NY, USA, 2023. Association for Computing Machinery.

[170] M. Van Mechelen, R. C. Smith, M.-M. Schaper, M. Tamashiro, K.-E. Bilstrup, M. Lunding, M. Graves Petersen, and O. Sejer Iversen. Emerging technologies in k–12 education: A future hci research agenda. *ACM Trans. Comput.-Hum. Interact.*, 30(3), jun 2023.

[171] J. Vincent. Twitter taught microsoft's ai chatbot to be a racist asshole in less than a day, Mar 2016.

[172] L. S. Vygotsky. *Thought and Language, revised and expanded edition*. MIT Press, July 2012.

[173] L. S. Vygotsky and M. Cole. *Mind in society: Development of higher psychological processes*. Harvard university press, 1978.

[174] J. Walrack, B. Segal, and S. Snider. Where do i fall in the american economic class system?, 2023.

[175] G. Wang, K. Sun, A. Atabey, K. Pothong, G. C. Lin, J. Zhao, and J. Yip. Child-centred ai design: Definition, operation, and considerations. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2023.

[176] G. Wang, J. Zhao, M. Van Kleek, and N. Shadbolt. 'don't make assumptions about me!': Understanding children's perception of datafication online. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–24, 2022.

[177] G. Wang, J. Zhao, M. Van Kleek, and N. Shadbolt. '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*, pages 1–21, 2023.

[178] L. Wang, M. I. Mujib, J. Williams, G. Demiris, and J. Huh-Yoo. An evaluation of generative pre-training model-based therapy chatbot for caregivers. *arXiv preprint arXiv:2107.13115*, 2021.

[179] S. Wang, Y. Duan, H. Ding, Y.-P. Tan, K.-H. Yap, and J. Yuan. Learning transferable human-object interaction detector with natural language supervision. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[180] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT, Feb. 2023. arXiv:2302.11382 [cs].

[181] G. Wiggins and J. McTighe. *Understanding by design*. Ascd, 2005.

[182] R. Williams, H. W. Park, and C. Breazeal. 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*, CHI '19, page 1–11, New York, NY, USA, 2019. Association for Computing Machinery.

[183] J. Wirth, J. Künsting, and D. Leutner. The impact of goal specificity and goal type on learning outcome and cognitive load. *Computers in Human Behavior*, 25(2):299–305, 2009. Including the Special Issue: State of the Art Research into Cognitive Load Theory.

[184] T. Wu, M. Terry, and C. J. Cai. 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, 2022.

[185] W. Xu, J. Ma, J. Yao, W. Lin, C. Zhang, X. Xia, N. Zhuang, S. Weng, X. Xie, S. Feng, F. Ying, P. Hansen, and C. Yao. Mathkingdom: Teaching children mathematical language through speaking at home via a voice-guided game. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[186] Y. Xu, S. Branham, X. Deng, P. Collins, and M. Warschauer. 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, New York, NY, USA, 2021. Association for Computing Machinery.

[187] Y. Xu, S. Branham, X. Deng, P. Collins, and M. Warschauer. 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, pages 1–12, New York, NY, USA, 2021. Association for Computing Machinery.

[188] Y. Xu, S. Branham, X. Deng, P. Collins, and M. Warschauer. Are current voice interfaces designed to support children's language development? In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–12, 2021.

[189] Y. Xu, K. He, V. Vigil, S. Ojeda-Ramirez, X. Liu, J. Levine, K. Cervera, and M. W. Uci. "Rosita Reads With My Family": Developing A Bilingual Conversational Agent to Support Parent-Child Shared Reading, Feb. 2023.

[190] Y. Xu, K. He, V. Vigil, S. Ojeda-Ramirez, X. Liu, J. Levine, K. Cervera, and M. Warschauer. "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*, IDC '23, page 160–172, New York, NY, USA, 2023. Association for Computing Machinery.

[191] Y. Xu, V. Vigil, A. S. Bustamante, and M. Warschauer. "Elinor's Talking to Me!":Integrating Conversational AI into Children's Narrative Science Programming. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–16, New York, NY, USA, Apr. 2022. Association for Computing Machinery.

[192] Y. Xu and M. Warschauer. 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, pages 1–8, New York, NY, USA, May 2019. Association for Computing Machinery.

[193] Y. Xu and M. Warschauer. A content analysis of voice-based apps on the market for early literacy development. In *Proceedings of the Interaction Design and Children Conference*, IDC '20, pages 361–371, New York, NY, USA, June 2020. Association for Computing Machinery.

[194] Y. Xu and M. Warschauer. 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*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[195] O. N. Yalçin, S. Lalle, and C. Conati. An intelligent pedagogical agent to foster computational thinking in open-ended game design activities. In *27th International Conference on Intelligent User Interfaces*, IUI '22, page 633–645, New York, NY, USA, 2022. Association for Computing Machinery.

[196] D. Yang, Y. Zhou, Z. Zhang, and T. J.-J. Li. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. *ACM IUI Workshop*, 2022.

[197] Q. Yang, A. Steinfeld, C. Rosé, and J. Zimmerman. 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, pages 1–13, New York, NY, USA, Apr. 2020. Association for Computing Machinery.

[198] A. Yuan, A. Coenen, E. Reif, and D. Ippolito. Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces*, IUI '22, pages 841–852, New York, NY, USA, 2022. Association for Computing Machinery.

[199] C. Zhang, X. Liu, K. Ziska, S. Jeon, C.-L. Yu, and Y. Xu. Mathemyths: Leveraging Large Language Models to Teach Mathematical Language through Child-AI Co-Creative Storytelling, Feb. 2024. arXiv:2402.01927 [cs].

[200] C. Zhang, C. Yao, J. Liu, Z. Zhou, W. Zhang, L. Liu, F. Ying, Y. Zhao, and G. Wang. 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, pages 1–6, New York, NY, USA, 2021. Association for Computing Machinery.

[201] Z. Zhang, Y. Xu, Y. Wang, B. Yao, D. Ritchie, T. Wu, M. Yu, D. Wang, and T. J.-J. Li. 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, pages 1–21, New York, NY, USA, 2022. Association for Computing Machinery.

[202] H. Zhou, M. Huang, T. Zhang, X. Zhu, and B. Liu. Emotional chatting machine: Emotional conversation generation with internal and external memory. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr. 2018.

# Appendix A

# Appendix Title

## A.1 Semi-structured interview protocols of the usability study with StoryAI for students

1. Story satisfaction (i.e., Do you like your story today?)
2. Efficacy of StoryAI (i.e., Was StoryAI helpful? why?)
3. Perception (Efficacy of StoryAI) (i.e., Would you like to write with StoryAI or without StoryAI)?
4. Learning (i.e, What do you think you learned from StoryAI).
5. Ownership (i.e., Do you think your story is yours or StoryAI's?).
6. Perception (AI's role) (i.e., What do you think of AI's role in your experience such as teachers/ peers)?

## A.2 Post survey protocols of the usability study with StoryAI for students

1. Ownership: I feel ownership over the final story.
2. Enjoyment: I enjoyed writing the story with StoryAI.
3. Ease: I found it easy to write the story with StoryAI.
4. Collaboration: I felt like I was collaborating with the AI in StoryAI.
5. Satisfaction: I'm proud of the final story.