Beyond "Just" Text: Can an AI-Generated Graphic Novel Enhance the Reading Experience of Non-Native English Readers?

Alice Vitali
Delft University of Technology
Delft, Netherlands
a.vitali@tudelft.nl

Isaac Alpizar-Chacon
Utrecht University
Utrecht, Netherlands
i.alpizarchacon@uu.nl

Christina Schneegass
Delft University of Technology
Delft, Netherlands
c.schneegass@tudelft.nl

Tilman Dingler
Delft University of Technology
Delft, Netherlands
t.dingler@tudelft.nl

Ioanna Lykourentzou
Utrecht University
Utrecht, Netherlands
i.lykourentzou@uu.nl

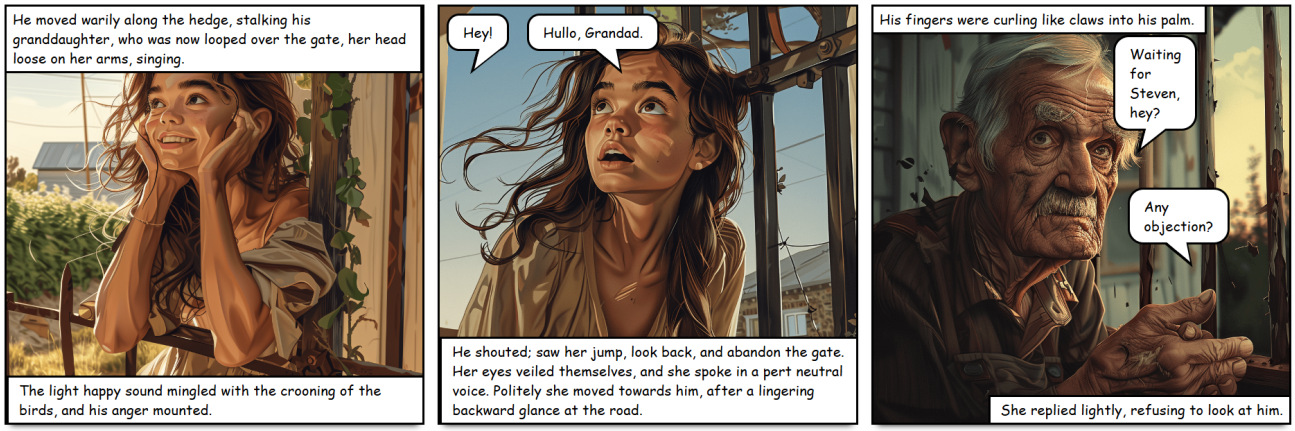


Figure 1: An excerpt from the final AI-generated graphic novel adaptation of Doris Lessing's short story "Flight", part of the collection "Habit of Loving" [39]. This graphic novel was first automatically generated using our proposed LangChain pipeline, and later used in the user study exploring ESL readers' perceptions of this AI-generated adaptation.

Abstract

Despite its well-documented benefits, reading literature in a second language remains challenging. Graphic novels, by integrating text and images, effectively support English as a Second language (ESL) readers; however, adapting literary works into this format is resource-intensive and lacks scalability. To address this, we developed a LangChain-based pipeline that automatically transforms a story into a graphic novel. Through a user study with 76 participants, we investigated (1) how this adaptation influences ESL readers' comprehension and narrative engagement, and (2) readers' perception of AI's role in the creative process. Results showed no significant differences in comprehension or engagement between the AI-generated graphic novel and traditional text. Although 70% of participants recognized AI involvement, attitudes toward its role as illustrator were generally positive, despite a few cross-domain concerns. This work contributes to the understanding of AI-powered storytelling from a human-centered perspective, identifying key insights for effectively supporting readers through AI-generated visual narratives.

CCS Concepts

• **Human-centered computing** → *Empirical studies in HCI*; • **Computing methodologies** → *Artificial intelligence.*

Keywords

Generative AI, LangChain, Creativity, Graphic Novel, Pleasure Reading, ESL Reading, GPT, Midjourney, English as a Second Language, Prompt Engineering, Perception, Engagement, Comprehension

ACM Reference Format:
Alice Vitali, Isaac Alpizar-Chacon, Christina Schneegass, Tilman Dingler, and Ioanna Lykourentzou. 2025. Beyond "Just" Text: Can an AI-Generated Graphic Novel Enhance the Reading Experience of Non-Native English Readers?. In *Mensch und Computer 2025 (MuC '25), August 31-September 03, 2025, Chemnitz, Germany*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3743049.3743076

1 Introduction

Literature and technology have a rich history of mutual influence. On the one hand, many real-world technologies have been inspired by science fiction; for instance, "Star Trek" technologies partially inspired the design of the first mobile phone [29, 35]. On

MuC '25, August 31-September 03, 2025, Chemnitz, Germany

ACM ISBN 979-8-4007-1582-2/25/08
https://doi.org/10.1145/3743049.3743076

the other hand, technological development continuously changes the way readers experience literary works, through new platforms (e.g., e-books) and forms of engagement (e.g., interactive storytelling) [65]. As a natural progression, recent advances in the field of generative AI have unlocked innovative possibilities to transform literary works, supporting and enhancing the reading experience.

Reading narratives yields well-documented benefits, both literary (e.g., improved vocabulary and comprehension) and non-literary (e.g. improved empathy and social skills) [15, 30]. These are especially relevant for ESL (English as a Second Language) users, as avid readers tend to excel in second language acquisition [3, 7, 17, 62]. Extensive reading leads to expanded vocabulary knowledge, improved grammar usage, and increased reading motivation, creating a positive feedback loop that supports language acquisition.

However, reading in a foreign language is often frustrating, due to uncommon vocabulary and syntax compromising the mental visualization of what is being read [16]—disrupting the reading flow. This is where graphic novels come in handy: because they integrate text and visuals, they are increasingly recognized as effective tools to support ESL reading [57]. Visual panels simplify textual interpretation and foster strong empathetic connections with characters [57], ultimately leading to higher comprehension, engagement and motivation to read [1].

The manual adaptation of narratives into graphic novels is, however, a resource-intensive creative process that lacks scalability. But what if we could automate it, leveraging a combination of large language models (LLMs) and text-to-image (TTI) models, such as GPT-4 and Midjourney?

To answer this question, we built a prompting pipeline that automatically transforms a short story into a graphic novel. As illustrated by the excerpt in Figure 1, this pipeline generates images that align with the text in terms of character appearances, key settings, and major narrative events, suggesting that this approach can produce coherent, meaningful illustrations. Moreover, it can be applied to texts that might otherwise never be illustrated, making literature more accessible to a broader ESL audience.

Yet, beyond technical feasibility, other important questions remain: does reading this AI-generated graphic novel yield the same benefits of traditional graphic novels? How does it influence readers' comprehension and narrative engagement? And how do readers perceive the involvement of AI in the creation of these graphic novels? As little is known about the complex interaction between comprehension, narrative engagement and reader perceptions, this study aims to address this gap by investigating ESL readers' responses to an AI-generated graphic novel, providing insights into the interplay between objective reading outcomes and subjective perceptions.

Thus, this study aims to answer the following research questions:

- **RQ1: How does reading an AI-generated graphic novel influence ESL readers' comprehension and narrative engagement, compared to a traditional non-illustrated text?**
- **RQ2: How do readers perceive AI's role in the creation of the graphic novel, and how does this perception influence their reading experience?**
  - RQ2.1: What are readers' general attitudes towards generative AI, and how do these influence their opinions on its utility in this context?
  - RQ2.2: (How) can readers detect AI's involvement in the graphic novel creation without being told?
  - RQ2.3: Once explicitly told that AI was used, does this information change readers' initial enjoyment?

By developing this prompting pipeline and investigating the reading experience from a human-centered perspective, we take a step toward effectively scaling the augmentation of text with imagery, particularly for ESL readers. The contributions of this study are threefold. Firstly, we propose a novel AI-powered pipeline for augmenting text with imagery, offering a scalable approach to give new appeal to literary works through graphic novel adaptations. Secondly, we explore the complex relationship between AI-generated visual narratives and the ESL reading experience, finding that, while comprehension and engagement weren't significantly enhanced, readers enjoyed the graphic novel regardless of AI involvement—challenging prevalent theories of algorithm aversion. Finally, we highlight key factors that can guide future implementations of effective AI-generated visual material to support reading, such as text-image complementarity and transparent attribution practices.

2 Related Work

This section explores prior research on the benefits of pleasure reading for ESL readers, how illustrations can support it, and existing work on automated comic generation.

The Importance of Reading. Stories constitute one of the fundamental elements of humanity [30]; in fact, reading narratives is much more than just a form of play or escapism: it helps raising educational standards and increasing personal well-being and development. Moreover, on top of literacy-related benefits, from vocabulary knowledge to text comprehension, it increases general knowledge, social skills, other cultures' understanding and community participation. In recent years, reading has experienced a significant transformation, due to the advent of digital media [43], expanding beyond the conventional notion of reading printed books. "Multimodal approaches to reading" encompass a diverse range of genres, such as comics, graphic novels, and song lyrics, along with various digital formats that blend text with other media like films, games, and music [74].

Reading in a Second Language. Long-form reading is particularly relevant for ESL learners [7, 62]: strong reading habits correlate with higher achievement in a second language (L2) [3], with the amount of extensive reading emerging as the only significant predictor of gains in L2 proficiency scores [17]. However, uncommon vocabulary and syntax make reading in a foreign language more difficult than pleasant [16], leading to the so-called "beginners paradox" [8]: "learners need to read to gain vocabulary knowledge, but they need vocabulary knowledge in order to read".

The Power of Illustrations. A large body of research [14, 40, 41, 63, 78] has investigated the impact of illustrations on reading, showing that picture books positively affect various reading aspects: vocabulary expansion, reading comprehension, story recall, grammatical awareness, reading time, concentration, attention and reading motivation. While the majority of studies focuses on young readers, evidence suggests that illustrations facilitate adult reading as well [2]. However, if the passage is easy

or already familiar, adults may ignore pictures; if the passage is too complex, visuals may not help. According to [24], pictures' effectiveness highly depends on their level of abstractness: the more concrete and detailed the visual, the least effective it tends to be.

Altogether, this evidence supports the "Cognitive Theory of Multimedia Learning" [48], suggesting that people learn more deeply from words and pictures than from words alone due to cross-channel representation of the same stimulus. For ESL readers, the benefits of text and images integration are particularly beneficial, as images can act as "instructional scaffolding" [6] for those who do not master peculiar vocabulary or sentence structures.

The Graphic Novel Format. Positioned between novels and comics [34], the graphic novel format embodies the integration of text and visuals. The use of graphic novels, especially if adaptations of classic literature, has shown numerous benefits in educational settings. Not only do they simplify textual interpretation and promote positive reading habits, but they also engage students more effectively compared to traditional prose [71].

Moreover, graphic novels are increasingly recognized as effective tools to support ESL reading, as visual panels help contextualize language within visual cues. Additionally, graphic novels' dynamic visual storytelling techniques appeal to ESL readers, encouraging them to actively engage with the material [57]. Although there is need for further quantitative research on the topic, [1]'s experimental study, examining the effect of reading a graphic novel on ESL reading comprehension, reaches the conclusion that this format has a significant positive effect on intermediate school students' reading comprehension.

Generative AI for Graphic Novel Creation. State-of-the-art generative models, such as large language models and text-to-image models, can now produce high-quality text and visuals with minimal human supervision, driving advances in AI-powered visual storytelling and automated comics generation.

Although previous research has explored the transformation of text [58, 82], videos [75], charts [80], and even code [69] into comics, these studies primarily focus on technical aspects, such as automated panel layout or character consistency. The potential of AI-generated graphic novels specifically for ESL readers represents a promising yet underexplored intersection of educational technology and creative AI applications.

Additionally, unlike prior work, our study not only streamlines the graphic novel creation process, but also explores readers' perceptions of the resulting adaptation, disclosing AI's involvement in its creation. The integration of generative AI in creative domains is in fact controversial, as creativity is often considered an exclusively human capability. However, little research has examined attitudes toward AI in creative fields [56], or how disclosing AI involvement affects the perception of generated content—revealing "human favoritism, not AI aversion" [79]. Previous studies [31, 44] demonstrated how humans can not distinguish between human-made and AI-generated poetry and images. And yet, both [50] and [60] suggest that there is an anthropocentric bias towards AI-generated art, as people tend to consider it overall inferior to human-made art.

3 Implementation: From Wall-of-Text to Graphic Novel

Our pipeline adapts a short story into a graphic novel through a two-stage process: (1) two LangChain [1] prompt chains transform the story into individual text prompts; (2) these prompts become input to Midjourney, which generates the graphic novel panels. The chains' structure was inspired by a publicly available [2] LangChain implementation for transforming text into one image [4], expanded to achieve multiple scenes' illustrations.

Technological Framework. We used **GPT-4o** [3] as LLM (through the OpenAI [4]'s API) and **Midjourney V6** [5] as image generator. Both are closed-source, state-of-the-art generative models that do not require resource-intensive hardware for inference. OpenAI's API enables easy integration of their pre-trained models within the **LangChain** framework, allowing the building of prompt chains where the output of each prompt automatically feeds into the next prompt as input. Although Midjourney lacks an API, it facilitates maintaining consistency across (multiple) characters through the use of reference images [13].

Short Story Selection. We selected the short story "Flight", from Doris Lessing's collection "Habit of Loving" [39] for the following pragmatic reasons: (1) its detailed descriptions enable effective image generation with minimal human intervention; (2) its length (1905 words) minimizes reader fatigue; (3) limited cast and dialogues allow consistent, uncrowded panel generation; (4) polished syntax and vocabulary from a Nobel Prize-winning author present an appropriate challenge for ESL adult readers; (5) while Lessing is renowned, this story is lesser-known, minimizing the risk of participants' prior exposure.

Character-Chain. To achieve accurate illustrations for the four characters, we designed **Character-Chain** (see Figure 2, top block) as a two prompts sequence. The first prompt employs one-shot learning to identify characters and retrieve their names. The second uses role prompting ("You are an expert of prompt engineering for Midjourney") to generate descriptions for each character, incorporating both explicit details and information inferred from context (e.g., suitable clothing for a rural setting). These descriptions (see Appendix A.1) became input to Midjourney, resulting in the reference images used to ensure character consistency throughout the whole process. Since Midjourney outputs four images for each prompt, one per character was arbitrarily selected; see Figure 9 in Appendix A.2 for the reference images of the four characters.

Story-Chain. To transform the original text into detailed prompts for the visual panels, we designed **Story-Chain** (see Figure 2, bottom block) as a five prompts sequence. Since no prior work had established best practices for this prompting task, the final structure was developed through iterative refinements: (1) each occurrence of a character in the original text is substituted with its proper name, previously extracted by Character-Chain; (2) the resulting text is divided into 50 chunks, corresponding to the main scenes; (3) each chunk is turned into an illustration description featuring a subject, the subject's details, and the environment description; lastly, (4) each description is transformed

[1] LangChain, https://www.langchain.com/
[2] LangChain RAG, https://github.com/aambekar234/learning-notebooks/blob/main/llm-rag/langchain.ipynb
[3] Hello GPT-4o, https://openai.com/index/hello-gpt-4o/
[4] OpenAI, https://openai.com/
[5] Midjourney Model Versions, https://docs.midjourney.com/docs/model-versions

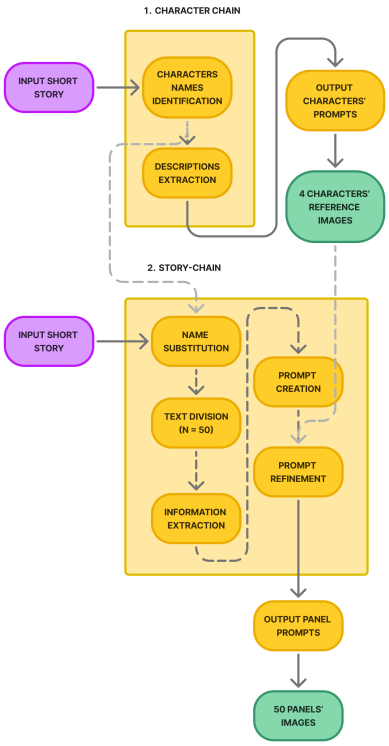


Figure 2: Prompting pipeline flowchart illustrating how the input story (purple) is processed sequentially. First, the Character-Chain transforms the text into character image prompts. Then, the Story-Chain generates prompts for story panels. Both sets of prompts are fed into Midjourney for image generation (green).

into a prompt for Midjourney, and (5) constant Midjourney parameters are added to the prompt (e.g., digital art, realism, graphic novel, —cref, —sref). See Appendix B for further details. The 49 output prompts became input to Midjourney, resulting in the graphic novel panels.

Graphic Novel Layout. Upon Midjourney's generation of the images, we created a digital graphic novel responsive layout using HTML and CSS. Adhering to the conventions of the graphic novel genre, we maintained core elements such as a panel layout and speech bubbles [34]; alternative formats, like slideshows, were initially considered but were dismissed as they would not capture the essence of the genre. The skeleton of the graphic novel layout is adapted from the CodePen project "Responsive Comic Book Layout" [6].

4 Experimental Methods

An ethical assessment instrument at [anonymous] was completed, classifying this user study as low-risk and deeming no additional ethics review necessary.

4.1 Study Design

We employed a pre-test/post-test control group design to address our research questions. Participants reading the AI-illustrated graphic novel formed the treatment group, while those reading the text-only version formed the control group.

(1) Pre-test — An initial English proficiency assessment ensured comparable English baselines between groups.
(2) Reading Intervention — Both groups read the short story "Flight". The treatment group read the AI-generated graphic novel, while the control group read the unillustrated version. The text was identical, except for minor rephrasings to fit speech bubbles.
(3) Post-test — Immediately after reading, participants of both groups completed a post-test questionnaire to assess comprehension and narrative engagement.
(4) Post-experimental disclosure — The treatment group completed an additional section in the post-test to assess their perception of AI in this context.

This experimental design isolates the effect of reading format while ensuring group comparability. Random assignment controlled for individual differences, in alignment with established methodologies for evaluating educational interventions [10, 23, 68, 70, 76, 77].

4.2 Participants

Over a period of four weeks, we recruited 90 participants through Prolific[7], 10 for the pilot study and 80 for the primary data collection. All participants provided informed consent and were compensated at an hourly rate of £9, in line with Prolific's fair pay guidelines. Prolific's pre-screeners ensured all participants met the following requirements: (1) any language other than English as their primary and earliest language in life; (2) raised monolingual (to exclude those raised bilingual in English); (3) fluent in English; (4) no language-related disorders or literacy difficulties. These prescreening criteria guaranteed the successful recruitment of a pool of adult ESL proficient readers. Moreover, we excluded participants who (1) failed two of the three attention checks placed within the survey, (2) consistently gave the same answer to all questions, and/or (3) reported prior familiarity with the story *Flight*. The three attention checks—two subtle Infrequency Items (IFIs) and one Instructed Response Item (IRI) [53]—helped filter out inattentive participants, ensuring data quality.

A pilot study was conducted to evaluate the survey flow, timing, and clarity of item wording, revealing no issues. The pilot participants (N = 10) were excluded from the main analysis. Of the remaining 80 participants, four were excluded for failing two of the three attention checks, resulting in a final sample of 76 participants (39 in the treatment group, 37 in the control group). No additional exclusions were made, as no participants consistently selected the same response option (e.g., always choosing "N/A"), and all confirmed it was their first time reading "Flight" by Doris Lessing.

4.3 Measures

Pre-test. The pre-test questionnaire was crafted by selecting only the items relevant to reading from two existing self-assessment measures: the CEFR [20] self-assessment orientation tool [8], and the Language Experience and Proficiency Questionnaire [9] (LEAP-Q) [47] [36]. The final set of questions, refined through review by two other authors, included general information (Prolific ID,

[6]Responsive Comic Book Layout, https://codepen.io/chris22smith/pen/MyBBOe

[7]Prolific, https://www.prolific.com/
[8]Assessment grid English, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168045bb52
[9]LEAP-Q, https://bilingualism.northwestern.edu/leapq/.

gender, age, and native language), years of formal education, CEFR reading level, percentage of daily life conducted in English, and the frequency of daily English activities (e.g., watching TV/movies), measured on a 8-items, 5-point Likert scale ranging from Never to Always.

Post-test. The post-test survey was designed to measure participants' comprehension, narrative engagement, and enjoyment. To measure participant's **reading comprehension**, we administered a comprehension test with 20 multiple-choice questions, crafted specifically for the narrative "Flight" by the author and refined through the revision of two other authors. We aimed to assess two levels of comprehension [64]: literal ("about what is explicitly stated") and inferential ("about what can be inferred but is not written"). Each correct answer was counted as 1 point and each incorrect answer was counted as zero points.

To measure participant's **narrative engagement**, we adopted the already existing Narrative Engagement Measure (NEM) [9, 21]—a validated self-reported measure with good reliability estimates which has already been applied to different narrative formats, "from TV shows to text-based stories, including short stories and comics" [9]. The NEM questionnaire measures narrative engagement through a 12-item, 5-point Likert scale, with scores averaged to provide an overall measure of engagement [21].

The treatment group answered two additional 5-point Likert scales. The first assessed their **enjoyment of the integration of text and images** through four items (enjoyment of text, enjoyment of images, how well images captured each scene's essence, and overall quality of text-image integration). The second assessed their **general enjoyment** through two items: willingness to continue reading (i.e., future intent) and overall enjoyment (i.e., immediate response) [38]. These ratings were collected prior to disclosing the role of AI, after which participants will be asked if they wish to revise their responses, allowing us to measure any significant changes in their ratings.

Post-experimental disclosure. The treatment group participants were explicitly informed about generative AI's role in creating the graphic novel. Following this disclosure, participants indicated whether they had suspected so and why; although asking the question after revealing AI's role may have biased their response, we believe that asking it outright, without prior context, could have also influenced participants by prompting them to consider AI as a possibility even if they initially had not. Next, they filled in four Likert scales measuring **attitudes** towards AI (13 items), **beliefs** about generative AI (seven items), **familiarity** with it (four items), and **perceived utility** of using AI as in this study (two items). Lastly, they were given the chance to change their initial **enjoyment** ratings in light of the AI disclosure and leave any optional open comments.

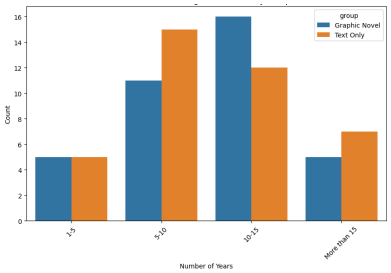
5 Data Analysis and Results

All steps of the data analysis were performed using Python version 3.11.3. We used Python libraries, including Pandas [10] for data manipulation, Scipy.stats [11] for statistical functions, Statsmodels [12] for statistical modeling, and Matplotlib [13] and Seaborn [14] for data visualization.

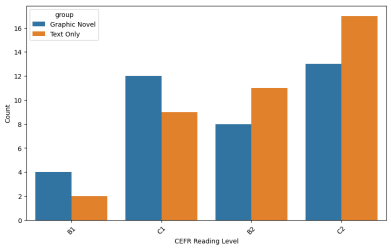
[10] Pandas, https://pandas.pydata.org/pandas-docs/stable/index.html
[11] Scipy.stats, https://docs.scipy.org/doc/scipy/reference/stats.html
[12] Statsmodel, https://www.statsmodels.org/stable/index.html
[13] Matplotlib, https://matplotlib.org/
[14] Seaborn, https://seaborn.pydata.org/

5.1 Pre-Test

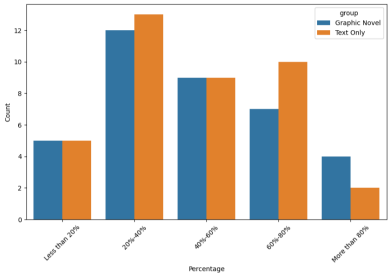
As shown in Figure 3, most treatment group participants had 10–15 years of formal English education, while control group participants had 5–10 years. The majority in both groups self-assessed their reading level as C2 (CEFR) and reported using English 20–40% of the time in daily life. In terms of specific daily English activities, the two groups show mostly similar patterns (see Figure 4). Listening to music or podcasts in English is the most common activity, followed by watching TV shows, movies, or videos, then using social media, and finally reading the news. On the other hand, participants rarely communicate in English with their families. However, most treatment group participants reported reading seldom for both academic and pleasure purposes, whereas most control group participants reported reading often. **No statistically significant difference** was observed across any of the dimensions measured by the pre-test (all $p > 0.05$); this indicates that the treatment and control groups were comparable in terms of baseline English proficiency, validating the subsequent analysis of the intervention effects.



(a) Years of Formal Education



(b) CEFR Reading Levels



(c) Percentage of Daily English Use

Figure 3: Bar plots comparing the treatment group ("Graphic Novel") and control group ("Text Only") across three pre-test variables: years of formal english education, CEFR reading level and percentage of daily english use.

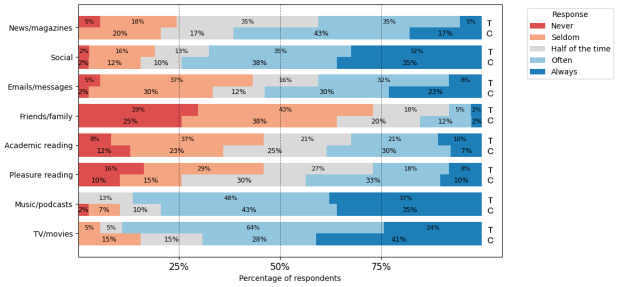


Figure 4: Stacked bar plot comparing daily engagement frequency across 8 English activities for the treatment (T) and control (C) groups. Each activity shows T above C, with color-coded bars indicating response distributions from 'Never' (red) to 'Always' (dark blue).

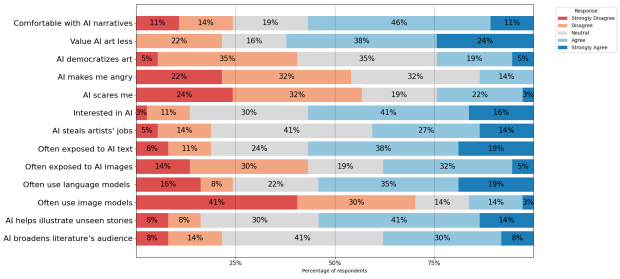


Figure 6: Stacked bar plot summarizing participants' responses to a 13-item, 5-point Likert scale measuring beliefs about AI (first 7 items), familiarity with AI tools/content (next 4 items), and perceived utility of AI in this study (last 2 items).

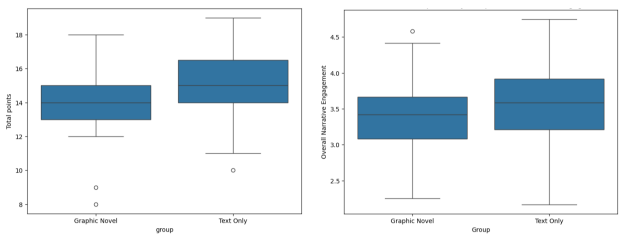
5.2 Post-Test

Comprehension. The treatment group scored on average 14.3/20 (SD = 2.2) on the comprehension test, while the control group's average was 14.9/20 (SD = 2.0). Refer to Figure 5a for a visual comparison of the two groups. To address RQ1, we compared the two groups' comprehension scores with a t-test, finding **no statistically significant difference in comprehension accuracy (p = 0.184)**.

Narrative Engagement. The average narrative engagement was 3.42 (SD = 0.54) for the treatment group and 3.44 for the control group (SD = 0.74). See Figure 5b for a visual comparison of the two groups. To address RQ1, we compared the two groups' narrative engagement scores with a t-test, finding **no significant difference between groups in narrative engagement (p = 0.344)**.

5.3 Post-experimental Disclosure

Opinions, Beliefs and Familiarity. To address RQ2.1, we first analyzed participants' *beliefs about AI* (first 7 items), *familiarity with AI tools/content* (next 4 items) and *perceived utility of using AI as in this study* (last 2 items), shown in Figure 6. Our findings suggest that most participants are comfortable with the use of AI to illustrate narratives (57%), and agree that AI could illustrate stories that might otherwise never be illustrated (55%), potentially making literature more appealing to a broader audience (38%); yet, AI's potential to democratize art is met with mixed reactions. Emotion-related responses suggest a generally positive outlook on AI, with participants being more interested in it than angry or scared. However, skepticism about the value of



(a) Comprehension Accuracy **(b) Narrative engagement**

Figure 5: Box plots comparing the treatment group ("Graphic Novel") and control group ("Text Only") on comprehension and narrative engagement.

AI-generated illustrations is evident, with 62% of the participants believing that they have less value than human-crafted ones. Concerns about AI's impact on artistic job displacement also emerge. Participants reported higher exposure to AI-generated textual content (57%) than visual content (37%). This trend is even more marked when referring to their use of text and image models: while 54% regularly uses GPT-like models, 71% claims not to use image generators.

To explore how participants' beliefs about/familiarity with AI (independent variables) predict their views on AI's utility in this context (dependent variables), we used **ordinal logistic regression**. Among the *opinion statements*, only one significantly predicted the first dependent variable, indicating that **participants who believe AI democratizes art are more likely to perceive AI as useful for illustrating stories that would otherwise not be illustrated** (p = 0.026; coefficient = 1.099). Similarly, only one *opinion statement* is a predictor of the second dependent variable, indicating that **participants who are comfortable with AI illustrating narratives are more likely to believe that AI can make literature more appealing to a broader audience** (p = 0.004; coefficient = 1.726). Lastly, only one of the *familiarity statements* was found to be a negative predictor of both dependent variables, indicating that **participants who are more often exposed to AI-generated visual content are less likely to perceive AI as useful for illustrating stories that would otherwise remain unillustrated** (p = 0.015; coefficient = -0.842), **and less likely to believe that AI can make literature more appealing to a broader audience** (p = 0.022; coefficient = -0.735).

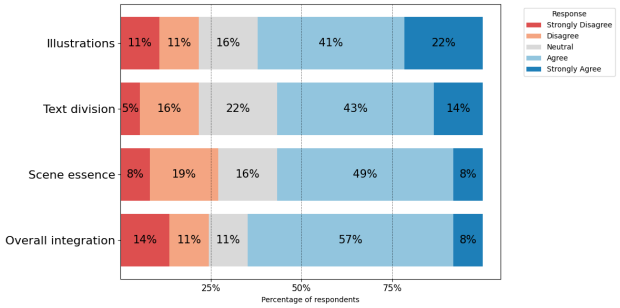
AI Involvement Detection. To answer RQ2.2, we first calculated the percentage of participants who suspected AI involvement, finding that 26 of the 37 participants (70.3%) did. We then analyzed their open-ended comments using a keyword-driven [67], **conceptual content analysis** [18, 67]; in other words, we combined computational techniques with qualitative interpretation to identify, group and quantify recurring themes in participants' comments. Specifically, (1) we pre-processed answers by expanding contractions, converting text to lowercase, and lemmatizing words. (2) We manually and iteratively extracted the main semantic concepts and their associated keywords (inductive approach [46]). (3) Based on these keywords, themes were assigned to each response via fuzzy string matching. Lastly, (4) we analyzed the frequency of each theme across participants' answers to understand the main reasons that led them to suspect AI's involvement in the

graphic novel creation. The following 4 main reasons emerged: (1) **General AI-"feel"** due to glossy, overly polished textures ("The illustrations seemed like AI-generated to me from the first glance."); (2) **Images' inconsistencies** in style and character details ("Some of them looked odd and sometimes the style would change just slightly, other times there were inconsistencies with e.g. the characters' hair or facial features."); (3) **General negative impressions**, with the graphic novel feeling "strange" and "lacking an individual artistic style" ("The illustrations kind of feel off [...] The style with a lot of close-ups also seemed weird."). (4) **Issues with text**, mainly related to speech bubbles and punctuation ("Also, the text panels looked a bit childish and as if they were made in paint and not by a graphic artist.") See Table 1 for the frequency of each concept and their associated keywords. These keywords were not generally linked to the themes beforehand but emerged inductively from the data; for example, *face* is connected to the theme "image inconsistencies" because it always appeared in relation to comments about visual incongruities. For visual examples of the uncanny facial features and clothing inconsistencies described by participants, refer to Figures 10a and 10b in Appendix C.

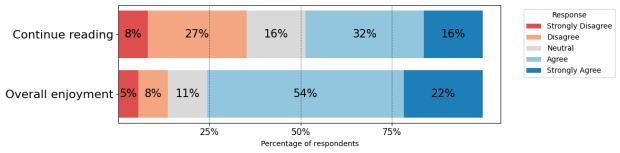
Enjoyment Change. Lastly, to answer RQ2.3, we analyzed the treatment group's enjoyment ratings, and how the AI disclosure impacted them. In general, participants of the treatment group responded positively, with the majority of ratings falling in the 'Agree' or 'Strongly Agree' range across both the text-image integration items and the more general enjoyment measures (see Figures 7a and 7b). Only 9 out of 37 participants (24.3%) changed their enjoyment ratings after learning about AI's role. Figure 8 shows a side-by-side comparison of their pre- and post-disclosure ratings. Since the data did not meet the assumptions for a paired t-test, we used a Wilcoxon signed-rank test to compare pre- and post-disclosure ratings. Even among those who adjusted their ratings, **no statistically significant difference was found in any of the 6 enjoyment-related items ($p > 0.05$ for all comparisons)**.

Concept	Occurrence	Concept's Keywords
General AI 'feel'	21	ai, quality, general, vivid, style, texture, graphic, illustration, figure
Images 'inconsistencies'	11	repeated, consistent, inconsistent, change, not similar, inconsistency, different, face
Negative impressions	6	hard, weird, feel off, strange, lack, odd
Issues with text	3	text, punctuation

Table 1: Summary of the concepts extracted from participants' open-ended comments through inductive, keyword-based content analysis. For each concept, the table shows the number of occurrences and its related keywords. Because comments could relate to multiple concepts, total frequencies exceed the number of comments.

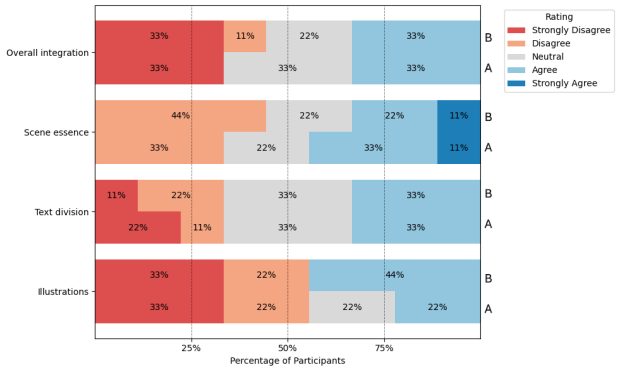


(a) Enjoyment of the integration of text and images, measured across four items (overall integration, scene essence, text division, and illustrations.

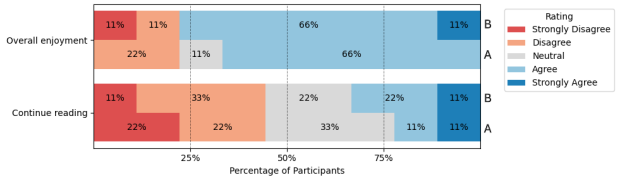


(b) General enjoyment of the graphic novel, based on two items (willingness to continue reading and overall enjoyment).

Figure 7: Stacked bar plots summarizing enjoyment ratings from all participants (N = 37) across (a) text-image integration and (b) general enjoyment.



(a) Change in enjoyment of text and image integration across four items (overall integration, scene essence, text division, and illustrations.



(b) Change in general enjoyment of the graphic novel, based on two items (willingness to continue reading and overall enjoyment).

Figure 8: Stacked bar plots showing enjoyment ratings before and after the AI disclosure, for participants who changed at least one rating (n = 9). For each item, the top bar shows responses before (B) AI disclosure, and the bottom bar those after (A).

End-of-survey Open Comments. Participants' open-ended comments can offering additional qualitative insights, complementing

this result section. Although we did not analyze them systematically, several echo earlier findings: some participants described the images as kitsch, hyper-realistic, inconsistent, or with exaggerated facial expressions. Positive remarks also emerged: some found the illustrations detailed, well-executed, or even entertaining. One praised their overall quality, while another expressed surprise at what generative models could produce (P: "I kinda felt strange cause i didnt even thought for a second that AI could be involved in it"). Only one participant raised concerns about the lack of legislation on generative AI. Another noted that the story was hard to read.

Those in the text-only condition largely offered positive remarks, expressing interest in the study and appreciation for discovering Doris Lessing. One participant found the story difficult. Feedback from the graphic novel group was more varied. While most were enthusiastic, others pointed out issues such as poorly placed text, overly complex vocabulary, and images failing to reflect key narrative element. One participant articulated a key limitation of the visual adaptation, noting that while the illustrations effectively presented the text visually, they failed in their complementary role, merely "doubling the same information presented via text", rather than highlighting key narrative details and emotional moments.

Sensitivity analysis. A sensitivity analysis ($\alpha = 0.05$, power = 0.80, $N_1 = 39$, $N_2 = 37$) conducted using G*Power [26] indicated that the study was powered to detect between-group differences of $d \geq 0.576$. Thus, while no statistically significant differences were found in several measures, smaller—but potentially meaningful—effects may have gone undetected.

6 Discussion

6.1 Comprehension and Narrative Engagement

To answer our first research question, we explored the impact of reading an AI-generated graphic novel on ESL participants' comprehension and narrative engagement. Results showed no significant differences compared to reading a text-only format, suggesting that the generated graphic novel did not measurably enhance nor hinder both aspects of the reading experience.

According to previous research [6, 42, 59], human-crafted illustrated adaptations of literary texts effectively enhance readers' comprehension [1, 19, 71] and engagement [14, 28, 42, 57, 59, 71]. In traditional graphic novels, however, each detail is carefully curated by expert writers and illustrators, as "visual panels are the text" [49]: colors, shading, panel layout, perspective, and lettering style are all integral parts of the story's meaning [66]. While we used role prompting to guide the generation, the generative models we used were not fine-tuned on the graphic novel genre, potentially limiting their ability to apply format-specific storytelling techniques.

Moreover, the effectiveness of illustrations highly depends on which, and how much, of the textual information they depict [11]. Despite the overall positive rating of integrating text and images, participants' qualitative feedback suggested that the images often mirrored, rather than complement, the text. For example, frequent close-ups of characters' faces may have limited the visuals' ability to highlight other key narrative details. As one participant aptly noted: "I felt like the illustrations often did a fine job of presenting the text in a visual manner, but that is not the sole purpose of illustrations in visual novels (in my opinion). I think they should complement the text, highlighting certain details and emotional beats, rather than just double the same information presented via text.".

Another factor explaining our results may be text complexity, as traditional graphic novels can exhibit different degrees of text simplification [28, 71]. In this study, the original text only underwent minor form modifications to ensure experimental rigor: should there be a notable difference between the treatment and control groups, we could attribute it with certainty to the visual panels and the smaller, more digestible, text chunks. Therefore, the lack of significant findings might also relate to the minimal text simplification applied, whereas other studies [1] may have utilized more simplified texts. Additionally, prior research states that if the passage is too complex, visuals may not help [2]; and indeed, participants highlighted difficulties with complex vocabulary and syntax, further supporting this interpretation: "Language was sometimes too hard for me - words which I don't understand were present".

These considerations may be particularly relevant for ESL readers, who face additional challenges when processing unfamiliar vocabulary and complex syntax. In such cases, the potential of visuals lies in strategically scaffolding difficult passages—either by depicting abstract or rare concepts or by reinforcing simplified textual chunks. Altogether, this underscores the need to explore how targeted alignment between text difficulty and image function might better support ESL comprehension.

6.2 AI Perceptions

To answer our second research question ("How do readers perceive AI's role in the creation of the graphic novel, and how does this perception influence their reading experience?"), we examined participants' attitudes toward AI-generated content and their ability to detect AI involvement.

6.2.1 Opinions, Beliefs and Familiarity. Our findings show that participants have an overall positive attitude towards the use of AI to illustrate narratives, as they recognize the potential to reach a broader audience by illustrating narratives that would otherwise remain unillustrated. However, our results align with previous findings [51, 61], reinforcing the notion that AI-generated art is perceived as inferior to human-made art— referred to as "anthropocentric bias".

Moving to trends in familiarity and exposure, our findings reflect insights from a recent Reuters Institute report [27]: while most people regularly use AI text generators (e.g., ChatGPT), they engage less frequently with AI image generators (e.g., DALL-E, Midjourney). Similarly, our participants reported more frequent exposure to AI-generated text than images. Interestingly, participants with more frequent exposure to generated images were less likely to perceive AI as useful in this context. This suggests that frequent exposure may lead to a more critical—if not skeptical—perspective, possibly due to having higher expectations or a better ability to detect inconsistencies. Such skepticism may also reflect broader societal effects of generative AI, which enables the mass creation of deepfakes—synthetic images or videos manipulated to appear real. Repeated exposure to deepfakes has been shown to foster uncertainty and reduce trust in digital content, particularly in online news contexts [73]; while our context differs, a similar mechanism may be at play.

6.2.2 AI Involvement Detection. With approximately 70% of the participants declaring they suspected AI involvement by looking

at the images, our results contradict previous research findings, which reported that humans could not distinguish human-made and AI-generated artworks [32, 45]. However, as the biggest issue with image generators is character consistency, which hardly becomes visible in standalone works of art, it's likely that the graphic novel format made AI detection easier—which is confirmed by the thematic analysis of participants' open comments. We could have manually smoothened out these inconsistencies, but since the aim was to streamline the graphic novel process, we consciously chose to keep human intervention minimal; additionally, we wanted to understand whether people would notice such image inconsistencies during reading, which can inform future research. Nevertheless, as generative AI continues to advance, we are confident that future iterations will achieve greater consistency and artistic fidelity. Even more than inconsistencies, participants mentioned a "general AI feel" due to the glossy and hyper-realistic images, which felt uncanny or even scary. A more traditional style could have worked better, resulting in a less striking AI presence.

6.2.3 Enjoyment Change. Since AI-generated artworks tend to be valued less than human-crafted ones [51, 61], and there is evidence for human favoritism in content creation [79], we could expect a decrease in enjoyment after disclosing AI usage. Instead, participants' enjoyment ratings did not change significantly. A closer look at the data reveals an interesting nuance: 8 of the 9 participants who chose to revise their ratings had already suspected AI's involvement. This suggests that AI disclosure did not impact enjoyment regardless of prior suspicion: those who already suspected AI were more likely to adjust their ratings when their suspicions were confirmed, but not significantly; those who did not initially spot AI chose not to change their ratings at all. This challenges the expectation that AI-generated content is inherently less enjoyable once its origin is revealed, in contrast to both the "human favoritism, not AI aversion" paradigm [79] and the "algorithm aversion" literature [12, 22].

6.3 Pulling the Strings Together: Practical Insights

The interplay between comprehension, engagement, and perceptions of AI reveals that, while visuals can support comprehension and engagement under ideal conditions [11], their effectiveness is not universal and depends on many factors. Simply adding visuals to text, without taking into account their complementary role or format- and genre-specific conventions, may not yield the intended benefits. Moreover, the high AI detection rate (70%) among participants raises important questions about how aesthetic perception influences the reading experience. Although the effect did not emerge from our quantitative data, qualitative insights—such as the "general AI feeling" and image inconsistencies—suggest its possible impact on narrative engagement. These observations inform the following practical insights for both supporting ESL readers and improving AI-generated graphic novel adaptations:

- Future AI-generated graphic novels should focus on creating visuals that complement rather than merely illustrate the text, perhaps through more sophisticated prompting techniques or fine-tuning models on graphic novel conventions. This is particularly important for non-native readers, as images can offer scaffolding to support a better understanding of the more challenging aspects of the text.
- More deliberate text simplification strategies might be necessary to fully leverage the benefits of visual support, particularly for ESL readers struggling with complex vocabulary and syntax.
- The "uncanny valley" effect of hyper-realistic AI art suggests that more traditional, stylized illustration approaches might be better received, at least in the context of literary materials.
- Given participants' positive attitude towards the use of AI to illustrate narratives and the lack of support for the algorithm aversion theory, generated material may benefit from transparent attribution practices.

Ethical Implications of Generative AI. Lastly, discussing the ethical implications of using generative models to transform texts into graphic novels is fundamental—especially since our system could, in principle, be scaled.

Concerns over artist job displacement emerged from some of our participants, in alignment with prior research [56]. While not a dominant theme, this highlights the importance of considering the stakeholders affected by generative AI—writers and illustrators, for instance [25]. Artists have reported unauthorized use of their work (e.g. closed-source training datasets using copyrighted content without consent) and exclusion from design and policy conversations [33]. In our specific context, we should also wonder to which extent writers should be involved in decisions about how their work is transformed, in order to respect their agency and artistic intent.

Moreover, questions of copyright, authorship and ownership remain unresolved [25]. Since both the user, the programmer and the model contribute to the output, it is unclear who should be credited or can profit from the results [81]. Currently, AI-generated images cannot be copyrighted [33]; however, exceptions exist: in February 2023, the US Copyright Office decided that a comic book created using AI ("Zarya of the Dawn" by [37]) was eligible for copyright protection to a certain extent — covering the written texts and the comic book as a whole but not the individual images [52] [55].

To avoid "claiming to create tools that democratize art without consulting artists" [5, 33, 72], these issues must be addressed before a pipeline like ours can be innocuously deployed.

Additionally, bias and discrimination remain critical concerns. Generative models tend to reproduce stereotypes embedded in their training data (e.g., racism, gender biases and western beauty standards), and may contribute to feedback loops that reinforce existing biases at the cost of diversity [25, 54]. This is already visible in our own results, as generated characters are all white, and the young girl and boy presented stereotypical attractive features. Privacy risks are also significant, as real images can be scraped and repurposed, enabling deepfakes and misinformation [73]. Unfortunately, regulation enforcement remains difficult, given the speed and decentralization of generative content production [33].

7 Limitations and Future Work

This study faced several technical and methodological limitations that should be addressed in future research. First, full automation was hindered by Midjourney's lack of an API, the need for human oversight to ensure character consistency, and the manual creation of the digital graphic novel's layout. However, ongoing advancements in generative AI will likely enable the full automation of these processes in the near future.

Although we used closed-source, proprietary AI models due to computational constraints, this choice involves trade-offs. Models like GPT and Midjourney cannot be fine-tuned to specific domains, likely limiting graphic novel quality by lacking genre-specific knowledge. Their opaque training data also raises ethical concerns, as mentioned above. Future work could improve outcomes by fine-tuning open-source models on graphic novel datasets, and integrating them into the same pipeline.

Secondly, the study's generalizability is limited by its focus on a single short story (Flight) and a single artistic style (hyper-realistic). While the pipeline demonstrated the feasibility of generating graphic novels from text, future work should investigate its scalability to longer works and diverse genres. In fact, the effectiveness of our proposed method is likely to vary across literary genres: while the method was not applied to other narratives, the prompting strategy was not tailored to the content specifically. Rather, it was genre-oriented, focusing on extracting generalizable elements such as characters, environments, and actions. This suggests that the pipeline may transfer to other similarly structured stories—narratives with clear actions, defined spatial relationships, and distinct characters—even though the final visuals would likely benefit from manual refinement.

Coming to the user study, the absence of standardized instruments for measuring AI perceptions led to the development of a custom questionnaire, underscoring the need for validated scales in future research.

Moreover, variations in graphic novel familiarity and preferences for visual versus textual information may have influenced comprehension, engagement, and overall perception of the generated graphic novel. Future studies could control for these factors through stratification or pre-testing to understand how these individual differences affect the reading experience.

8 Conclusion

This paper tapped into the unexplored potential of generative AI to streamline the adaptation of narratives into graphic novels, with the ultimate goal of supporting ESL reading. Through a human-centered approach, we investigated the interplay between comprehension, narrative engagement, and perceptions of AI's role in automatically adapting literary texts. Our results highlight both the potential and the challenges of such an approach, not only bringing us closer to the vision of effectively scaling the augmentation of text with imagery, but also informing the design of future AI-driven storytelling experiences. In conclusion, while our AI-generated graphic novel did not significantly enhance ESL readers' comprehension and engagement compared to a traditional text, our findings offer valuable insights into the complex interaction between AI-generated visual storytelling and reader perceptions. As generative AI technology continues to evolve rapidly, these insights can guide a more effective integration of AI-generated content in literary contexts.

References

[1] Reema Aldahash and Sultan Altalhab. 2020. The Effect of Graphic Novels on EFL Learners' Reading Comprehension. *International Journal of Applied Linguistics and English Literature* 9 (09 2020), 19. https://doi.org/10.7575/aiac.ijalel.v.9n.5p.19

[2] Kathryn L. Alesandrini. 1984. Pictures and Adult Learning. *Instructional Science* 13, 1 (1984), 63–77. http://www.jstor.org/stable/23369073

[3] Ahmed Alkilabi. 2015. The Place of Reading Comprehension in Second Language Acquisition. *Journal of Literature, Languages and Linguistics* 31 (02 2015), 1–23.

[4] Abhijeet Ambekar. 2024. Exploring the Application of Langchain: Transforming Text into Art with AI. https://www.linkedin.com/pulse/exploring-application-langchain-transforming-text-art-ambekar-jd3oc/ [Last accessed 22-01-2025].

[5] Arte es Ética. 2023. *Arte es Ética presente en la ACM FAccT Conference 2023.* https://arteesetica.org/arte-es-etica-presente-en-laacm-facct-conference-2023/ Accessed: 2023-06-03.

[6] Lottie Baker. 2015. How Many Words Is a Picture Worth? Integrating Visual Literacy in Language Learning with Photographs. *English Teaching Forum* 53, 4 (2015), 2–13.

[7] Barbara Buchorn-Stoll. 2002. The influence of storybook reading on language development. *Language Matters* 33, 1 (2002), 25–48. https://doi.org/10.1080/10228190208566178 arXiv:https://doi.org/10.1080/10228190208566178

[8] Miriam Burt, Joy Peyton, and Rebecca Adams. 2003. *Reading and Adult English Language Learners: A Review of the Research. Series on Preparing Adult English Language Learners for Success.* Technical Report. Center for Applied Linguistics, Washington, DC.; National Center for ESL Literacy Education, Washington, DC.

[9] Rick W. Busselle and Helena Bilandzic. 2009. Measuring Narrative Engagement. *Media Psychology* 12, 4 (Nov. 2009), 321–347. https://doi.org/10.1080/15213260903287259

[10] Fatma Bölükbas. 2013. The effect of reading strategies on reading comprehension in teaching Turkish as a foreign language. *Educational Research and Reviews* 8, 21 (2013), 2147.

[11] Russell N. Carney and Joel R. Levin. 2002. Pictorial Illustrations Still Improve Students' Learning From Text. *Educational Psychology Review* 14, 1 (2002), 5–26. http://www.jstor.org/stable/23363486

[12] Noah Castelo, Maarten W. Bos, and Donald R. Lehmann. 2019. Task-Dependent Algorithm Aversion. *Journal of Marketing Research* 56, 5 (2019), pp. 809–825. https://www.jstor.org/stable/26967271

[13] Xi Chen and Jeanhun Chung. 2024. A Study on Character Consistency Generated in Midjourney V6 Technology. *International Journal of Advanced Smart Convergence* 13, 2 (2024), 142–147.

[14] Yan Piaw Chua, Hwee Joo Tan, Siew Kuan Christine Wong, Velan Narayanasamy Sakthi, Huong Ying Lu, Vyapury Dharmalingam, and Syah Mohamed Noor Azzam. 2013. Effects of Book Illustrations on Reading. (2013), 516–524.

[15] Christina Clark and Kate Rumbold. 2006. *Reading for Pleasure: A Research Overview.* Technical Report. National Literacy Trust.

[16] Aimee Cole, Ariadne Brown, Christina Clark, and Irene Picton. 2022. *Children and young people's reading engagement in 2022: Continuing insight into the impact of the Covid-19 pandemic on reading.* Technical Report. National Literacy Trust.

[17] Molly F. Collins. 2010. ELL preschoolers' English vocabulary acquisition from storybook reading. *Early Childhood Research Quarterly* 25, 1 (2010), 84–97. https://doi.org/10.1016/j.ecresq.2009.07.009

[18] Columbia University Mailman School of Public Health. 2016. *Content Analysis Method and Examples.* https://www.publichealth.columbia.edu/research/population-health-methods/content-analysis Accessed: 2025-06-02.

[19] Mike P. Cook. 2016. Now I "See": The Impact of Graphic Novels on Reading Comprehension in High School English Classrooms. *Literacy Research and Instruction* 56, 1 (2016), 21–53. https://doi.org/10.1080/19388071.2016.1244869

[20] Council of Europe. 2001. *Common European framework of reference for languages.* Cambridge University Press, Cambridge, England.

[21] Jenny L. Crowley and Jennifer A. Jackl. 2017. Narrative Engagement Measure (NEM): (Busselle 'I&' Bilandzic, 2009). , 469–474 pages. https://doi.org/10.1002/9781119102991.ch51

[22] Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. 2015. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (Feb 2015), 114–126. https://doi.org/10.1037/xge0000033 Published online: 2014 Nov 17.

[23] Erhan Durukan. 2011. Effects of cooperative integrated reading and composition (CIRC) technique on reading-writing skills. *Educational Research and Reviews* 6, 1 (2011), 102.

[24] Francis M. Dwyer. 1978. *Strategies for Improving Visual Learning: A Handbook for the Effective Selection, Design, and Use of Visualized Materials.* Learning Services. https://books.google.nl/books?id=GG-7AAAAIAAJ

[25] Ziv Epstein, Aaron Hertzmann, the Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R. Frank, Matthew Groh, Laura Herman, Neil Leach, Robert Mahari, Alex Sandy Pentland, Olga Russakovsky, Hope Schroeder, and Amy Smith. 2023. Art and the science of generative AI. *Science* 380, 6650 (2023). https://doi.org/10.1126/science.adh4451

[26] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41, 4 (2009), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

[27] Richard Fletcher and K. Rasmus Nielsen. 2024. *What does the public in six countries think of generative AI in news?* Technical Report. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2024-05/Fletcher_and_Nielsen_Generative_AI_and_News_Audiences.pdf

[28] Dalila Forni. 2022. *Dai classici della letteratura ai graphic novel. Adattare i Classici a fumetti per una nuova educazione letteraria e visuale.* PensaMultimedia, Bergamo-Lecce, 247–270. https://www.unisob.na.it/ateneo/centrolll/volume2022_1.pdf#page=247

[29] Emerging Technology from the arXiv. 2018. When science fiction inspires real technology — technologyreview.com. https://www.technologyreview.

com/2018/04/05/67057/when-science-fiction-inspires-real-technology/. [Last accessed 22-01-2025].
[30] Elise N. Good and Katharine Schaab. 2022. The Biological Influence of Stories & The Importance of Reading Fiction. *The Kennesaw Journal of Undergraduate Research* 9 (2022). Issue 1. https://doi.org/10.62915/2474-4921.1234
[31] Jimpei Hitsuwari, Yoshiyuki Ueda, Woojin Yun, and Michio Nomura. 2023. Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior* 139 (2023), 107502. https://doi.org/10.1016/j.chb.2022.107502
[32] Jimpei Hitsuwari, Yoshiyuki Ueda, Woojin Yun, and Michio Nomura. 2023. Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior* 139 (2023), 107502. https://doi.org/10.1016/j.chb.2022.107502
[33] Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. AI Art and its Impact on Artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) *(AIES '23)*. Association for Computing Machinery, New York, NY, USA, 363–374. https://doi.org/10.1145/3600211.3604681
[34] Petr Johanes. 2018. The Graphic Novel: A Promising Medium for Learning Research. In *2018 ASEE Annual Conference & Exposition*. ASEE Conferences, Salt Lake City, Utah. https://peer.asee.org/31103.
[35] Philipp Jordan, Omar Mubin, Mohammad Obaid, and Paula Alexandra Silva. 2018. Exploring the Referral and Usage of Science Fiction in HCI Literature. In *Design, User Experience, and Usability: Designing Interactions*, Aaron Marcus and Wentao Wang (Eds.). Springer International Publishing, Cham, 19–38.
[36] M. Kaushanskaya, H. K. Blumenfeld, and V. Marian. 2020. The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition* 23, 5 (2020), 945–950. https://doi.org/10.1017/S1366728919000038
[37] K. Kristina. [n. d.]. Zarya Of the Dawn. https://aicomicbooks.com/wp-content/uploads/2022/09/kashtanova-full-book-v1-small.pdf
[38] Sari Kujala and Miron-Shatz Talya. 2015. The Evolving Role of Expectations in Long-Term User Experience. https://doi.org/10.1145/2818187.2818271
[39] Doris M. Lessing. 1981. *Habit of Loving*. Penguin Books, Harlow, England.
[40] Howard W. Levie and Richard Lentz. 1982. Effects of Text Illustrations: A Review of Research. *Educational Communication and Technology* 30, 4 (1982), 195–232. http://www.jstor.org/stable/30219845
[41] Joel R. Levin. 1979. *On Functions of Pictures in Prose*. Technical Report. Wisconsin Univ., Madison. Research and Development Center for Individualized Schooling.
[42] Sarah D. A. Lim. 2010. *Exploring the Use of Picture Books in Developing the Language Acquisition of Students Learning English as a Seecond Language*. Master's thesis. Master of Education, Dept. of Language and Literacy Education, University of British Columbia.
[43] Ziming Liu. 2005. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *J. Documentation* 61 (2005), 700–712. https://api.semanticscholar.org/CorpusID:29803567
[44] Zeyu Lu, Di Huang, Lei Bai, Jingjing Qu, Chengyue Wu, Xihui Liu, and Wanli Ouyang. 2023. Seeing is not always believing: Benchmarking Human and Model Perception of AI-Generated Images. arXiv:2304.13023 [cs.AI]
[45] Zeyu Lu, Di Huang, Lei Bai, Jingjing Qu, Chengyue Wu, Xihui Liu, and Wanli Ouyang. 2023. Seeing is not always believing: Benchmarking Human and Model Perception of AI-Generated Images. arXiv:2304.13023 [cs.AI] https://arxiv.org/abs/2304.13023 [Last accessed 22-01-2025].
[46] I. MacPherson and L. McKie. 2010. *Qualitative research in programme evaluation*. SAGE Publications Ltd, 454–477. https://doi.org/10.4135/9781446268247.n28
[47] V. Marian, H. K. Blumenfeld, and M. Kaushanskaya. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research* 50, 4 (Aug. 2007), 940–967. https://doi.org/10.1044/1092-4388(2007/067)
[48] R. E. Mayer. 2005. *Cognitive Theory of Multimedia Learning*. Cambridge University Press, 31–48.
[49] Lori McLaughlin and Ph.D Jodi Pilgrim. 2021. Graphic Options: A Comparison of Stories and Their Graphic Novel Counterparts. https://api.semanticscholar.org/CorpusID:260388785
[50] Kobe Millet, Florian Buehler, Guanzhong Du, and Michail D. Kokkoris. 2023. Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior* 143 (June 2023), 1–9. https://doi.org/10.1016/j.chb.2023.107707
[51] Kobe Millet, Florian Buehler, Guanzhong Du, and Michail D. Kokkoris. 2023. Defending humankind: Anthropocentric bias in the appreciation of AI art. *Comput. Hum. Behav.* 143, C (June 2023), 9 pages. https://doi.org/10.1016/j.chb.2023.107707
[52] N. Mion. 2023. Artificial intelligence and copyright: Who owns AI-generated images? https://www.novagraaf.com/en/insights/artificial-intelligence-and-copyright-who-owns-ai-generated-images. [Accessed 12-03-2024].
[53] M. Muszynski. 2023. Attention checks and how to use them: Review and practical recommendations. *Ask Research and Methods* 32 (12 2023), 3–38. https://doi.org/10.18061/ask.v32i1.0001
[54] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5 Accessed: 2025-06-03.
[55] United States Copyright Office. [n. d.]. https://www.copyright.gov/docs/zarya-of-the-dawn.pdf
[56] Jonas Oppenlaender, Aku Visuri, Ville Paananen, Rhema Linder, and Johanna Silvennoinen. 2023. Text-to-Image Generation: Perceptions and Realities. arXiv:2303.13530 [cs.HC] https://arxiv.org/abs/2303.13530 [Last accessed 22-01-2025].
[57] Maribel Ortega. 2020. *Engaging English Language Learners Through the Use of Graphic Novels*. Master's thesis. Capstone Projects and Master's Theses. https://digitalcommons.csumb.edu/caps_thes_all/837
[58] Ben Proven-Bessel, Zilong Zhao, and Lydia Chen. 2021. ComicGAN: Text-to-Comic Generative Adversarial Network. arXiv:2109.09120 [cs.CV] https://arxiv.org/abs/2109.09120 [Last accessed 22-01-2025].
[59] Purwanti Purwanti and Bachrudin Musthafa. 2019. The Role of Illustration in Enhancing Reading Comprehension of College Students. *Edusentris* 5 (03 2019), 133. https://doi.org/10.17509/edusentris.v5i3.300
[60] Martin Ragot, Nicolas Martin, and Salomé Cojean. 2020. AI-generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence?. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA,) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3334480.3382892
[61] Martin Ragot, Nicolas Martin, and Salomé Cojean. 2020. AI-generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence?. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3334480.3382892
[62] Willy A. Renandya, B. R. Sundara Rajan, and George M. Jacobs. 1999. Extensive Reading With Adult Learners of English as a Second Language. *Relc Journal* 30 (06 1999), 39–60. https://doi.org/10.1177/003368829903000103
[63] Jennifer Rusted and Veronika Coltheart. 1979. The effect of pictures on the retention of novel words and prose passages. *Journal of experimental child psychology* 28, 3 (1979), 516–524. https://doi.org/10.1016/0022-0965(79)90078-x
[64] Mahmood Saadatnia, Saeed Ketabi, and Mansoor Tavakoli. 2017. Levels of Reading Comprehension Across Text Types: A Comparison of Literal and Inferential Comprehension of Expository and Narrative Texts in Iranian EFL Learners. *Journal of Psycholinguistic Research* 46, 5 (2017), 1087–1099. https://doi.org/10.1007/s10936-017-9481-3
[65] Elisabeth G. Sampel. 2022. The Influence of Modern Technology in Literature — edsa-unair.com. https://www.edsa-unair.com/2022/04/the-influence-of-modern-technology-in.html. [Last accessed 22-01-2025].
[66] Gretchen Schwarz. 2006. Expanding Literacies through Graphic Novels. *English Journal* 95, 6 (2006), 58. https://doi.org/10.2307/30046629
[67] Clive Seale and Jonathan Charteris-Black. 2010. *Keyword analysis: a new tool for qualitative research*. SAGE Publications Ltd, 536–556. https://doi.org/10.4135/9781446268247.n28
[68] Agus Sholeh, Punaji Setyosari, Bambang Y. Cahyono, and Sulthoni Sulthoni. 2019. Effects of Scaffolded Voluntary Reading on EFL Students' Reading Comprehension. *International Journal of Instruction* 12, 4 (2019), 297–312.
[69] Sangho Suh, Jian Zhao, and Edith Law. 2022. CodeToon: Story Ideation, Auto Comic Generation, and Structure Mapping for Code-Driven Storytelling. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 16 pages. https://doi.org/10.1145/3526113.3545617
[70] Yasemin Sönmez and Süleyman E. Sulak. 2018. The effect of the thinking-aloud strategy on the reading comprehension skills of 4th grade primary school students. *Universal Journal of Educational Research* 6, 1 (2018).
[71] R. Thanya and Chandramohan Suganthan. 2023. *The Importance od Graphic Novels as Motivating Reading Among Young Learners*. Vol. 2. https://doi.org/10.25215/9392917082.21
[72] The Distributed AI Research Institute. 2023. *Stochastic Parrots Day: What's Next? A Call to Action*. https://peertube.dairinstitute.org/w/a2KCCzj2nyfrq5FvJbL9jY Accessed: 2023-07-06.
[73] Cristian Vaccari and Andrew Chadwick. 2020. Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society* 6, 1 (2020), 2056305120903408. https://doi.org/10.1177/2056305120903408 arXiv:https://doi.org/10.1177/2056305120903408
[74] Ana Vogrinčič Čepič, Tiziana Mascia, and Juli-Anna. Aerila. 2024. Reading for Pleasure: A Review of Current Research. *New Zealand Journal of Educational Studies* (03 2024). https://doi.org/10.1007/s40841-024-00313-x
[75] Xin Yang, Zongliang Ma, Letian Yu, Ying Cao, Baocai Yin, Xiaopeng Wei, Qiang Zhang, and Rynson W. H. Lau. 2021. Automatic Comic Generation with Stylistic Multi-page Layouts and Emotion-driven Text Balloon Generation. arXiv:2101.11111 [cs.CV] https://arxiv.org/abs/2101.11111 [Last accessed 22-01-2025].
[76] Mohammad M. Yazdani and Mojtaba Mohammadi. 2015. The explicit instruction of reading strategies: directed reading thinking activity vs. guided reading strategies. *International Journal of Applied Linguistics and English Literature* 4, 3 (2015), 53–60.

[77] Jie Yu, Xing Zhou, Xiaoming Yang, and Jie Hu and. 2022. Mobile-assisted or paper-based? The influence of the reading medium on the reading comprehension of English as a foreign language. *Computer Assisted Language Learning* 35, 1-2 (2022), 217–245. https://doi.org/10.1080/09588221.2021.2012200 arXiv:https://doi.org/10.1080/09588221.2021.2012200

[78] Arbresha Zenki-Dalipi and Rabije Osmani. 2022. The Positive Influence of Images and Graphic Presentations in Teaching Texts for Keeping Students' Attention and COncentration. *Knowledge - International Journal* 54 (09 2022), 751–756. https://doi.org/10.35120/kij5405751z

[79] Yunhao Zhang and Renée Gosline. 2023. Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. *Judgment and Decision Making* 18 (2023), e41. https://doi.org/10.1017/jdm.2023.37

[80] Jian Zhao, Shenyu Xu, Senthil Chandrasegaran, Chris Bryan, Fan Du, Aditi Mishra, Xin Qian, Yiran Li, and Kwan-Liu Ma. 2021. ChartStory: Automated Partitioning, Layout, and Captioning of Charts into Comic-Style Narratives. arXiv:2103.03996 [cs.HC] https://arxiv.org/abs/2103.03996 [Last accessed 22-01-2025].

[81] K.-Q. Zhou and H. Nabus. 2023. The Ethical Implications of DALL-E: Opportunities and Challenges. *Mesopotamian Journal of Computer Science* 2023 (2023), 17–23.

[82] Yupeng Zhou, Daquan Zhou, Ming-Ming Cheng, Jiashi Feng, and Qibin Hou. 2024. StoryDiffusion: Consistent Self-Attention for Long-Range Image and Video Generation. arXiv:2405.01434 [cs.CV] https://arxiv.org/abs/2405.01434 [Last accessed 22-01-2025].

A Character-Chain

A.1 Characters Descriptions

Below are the 4 output prompts for Midjourney returned by Character-Chain, one per character:

- **Grandad:** "Create an illustration of an elderly man with weathered skin and a stern expression, his hair white and thinning. He wears simple, worn clothing suitable for outdoor work, with sleeves rolled up and a hint of dirt on his garments, indicative of a life spent laboring on the land. The time is late afternoon, and the setting suggests a rural environment in the early 20th century. Realistic portrait, 8k, Nikon, color digital style."
- **Alice:** "Generate an image of a young woman with sunlit, wavy hair cascading down her back. She has a youthful, carefree appearance, with long bare legs and a light summer dress that reflects the warm end-of-summer month in a rural village setting of the early 20th century. Her attire is simple and unadorned, hinting at a life close to nature. Realistic portrait, 8k, Nikon, color digital style."
- **Steven:** "Illustrate a robust young man with a strong, physical build, suggestive of a postmaster's son from a rural village in the early 20th century. His hair is neatly combed, and he wears practical clothing that is clean but shows signs of manual work. His attire includes a button-up shirt with rolled sleeves and durable trousers. Realistic portrait, 8k, Nikon, color digital style."
- **Lucy:** "Depict a mature woman with calm eyes and a composed expression, her hair neatly done in a style befitting a woman of her status in the early 20th century rural setting. She is dressed in a simple, modest dress with a pattern appropriate for a day spent indoors, sewing or managing household affairs. The atmosphere suggests a serene late afternoon in a quiet, sunlit home. Realistic portrait, 8k, Nikon, color digital style."

A.2 Characters Images

Below are the four reference images generated by Midjourney from the prompts reported above and used as character references throughout the whole generation process.

(a) Granddad **(b) Alice**

(c) Steven **(d) Lucy**

Figure 9: The four reference images generated by Midjourney for the characters of the story: Granddad, Alice, Steven, and Lucy.

B Story-Chain

Below we detail each step of Story-Chain in greater details:

(1) **Name substitution** - The first prompt instructs the model to modify the original text by substituting each occurrence of a character with its proper name. This prompt leverages both the role prompting technique and the one-shot learning technique.
(2) **Division** - The model is then directly asked to partition the given text (output of the previous prompt) into 50 sections, making sure all sections have a similar amount of text. No prompt engineering technique is used here.
(3) **Storyboard** - The model is then asked to transform each scene into an illustration description featuring three elements: a subject (either human or not), the subject's details (e.g., expression, color) and the environment/setting (where the subject is located). Once again, the instructions include to avoid multiple characters when possible, not to skip any scene and to create dissimilar panels. This prompt leverages both the role prompting technique and the one-shot learning technique.
(4) **Prompts** - The fourth prompt instructs the models to convert each panel resulted from the division step into a suitable prompt tailored for Midjourney. Detailed instructions are given: assume the model knows the characters by name; add words to describe time of the day, mood and camera angle for the shot; avoid direct speech or sound descriptions; return a unique sentence per panel; do not skip any of the input scenes. This prompt also leverages the role prompting technique.
(5) **Parameters** - The fifth and last prompt of the chain aims to format the output generated by the previous prompt, so

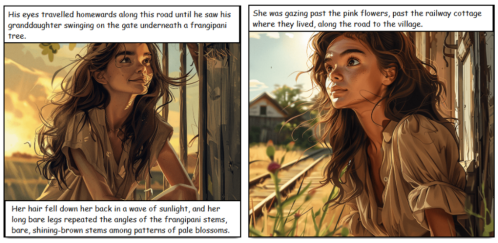
that they are suitable to be pasted directly into Midjourney. It does so by adding the string 'digital art, realism, graphic novel' and the parameter '–cref' followed by '–' and the name of the main human character present in the prompt (only one). Additionally, it adds the parameter '–sref', relevant for manually ensuring style consistency across panels.

Overall, this pipeline led to the generation of 50 prompts; only the last prompt, which contained simply 'The End,' was discarded. This number closely aligns with the target of 50 panels for the final graphic novel, ensuring a balanced distribution of text and visuals throughout the narrative.

C Examples of Uncanny and Inconsistent Images

The following images were selected to illustrate issues described by participants in their open-text responses, such as unsettling facial features and inconsistencies in character appearance. Although participants did not reference these specific images directly, their descriptions motivated the selection.

(a) Uncanny facial features reported by participants (e.g., exaggerated and scary facial expressions of the old man)



(b) Visual inconsistencies in character depiction (e.g., changes in Alice's clothing)

Figure 10: Examples of images selected by the first author to reflect visual concerns raised by participants, including uncanny facial expressions and inconsistent visual representations.