

CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars

Hua Xuan Qin
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, China
hxqin682@connect.hkust-gz.edu.cn

Shan Jin
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, China
sjin752@connect.hkust-gz.edu.cn

Ze Gao
Emerging Interdisciplinary Areas
Hong Kong University of Science and
Technology
Hong Kong, Hong Kong, China
zgaoap@connect.ust.hk

Mingming Fan
Computational Media and Arts Thrust
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, China
The Hong Kong University of Science
and Technology
Hong Kong, China
mingmingfan@ust.hk

Pan Hui
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, China
The Hong Kong University of Science
and Technology
Hong Kong, China
panhui@ust.hk

ABSTRACT

Support for story character construction is as essential as characters are for stories. Building upon past research on early character construction stages, we explore how conversation with chatbot avatars embodying characters powered by more recent technologies could support the entire character construction process for creative writing. Through a user study (N=14) with creative writers, we examine thinking and usage patterns of CharacterMeet, a prototype system allowing writers to progressively manifest characters through conversation while customizing context, character appearance, voice, and background image. We discover that CharacterMeet facilitates iterative character construction. Specifically, participants, including those with more linear usual approaches, alternated between writing and personalized exploration through visualization of ideas on CharacterMeet while visuals and audio enhanced immersion. Our findings support research on iterative creative processes and the growing potential of personalizable generative AI creativity support tools. We present design implications for leveraging chatbot avatars in the creative writing process.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; *Human computer interaction (HCI)*; • **Computing methodologies** → *Discourse, dialogue and pragmatics*.

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

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

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0330-0/24/05
<https://doi.org/10.1145/3613904.3642105>

KEYWORDS

Creativity Support, Writing Assistants, Creative Writing, Human-AI Collaboration, Large Language Models

ACM Reference Format:

Hua Xuan Qin, Shan Jin, Ze Gao, Mingming Fan, and Pan Hui. 2024. CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3613904.3642105>

1 INTRODUCTION

Characters are the souls of various forms of narrative, from traditional ones (e.g., novels and screenplays) to more modern ones (e.g., movies, computer games, and virtual museum tours [104, 117]). Support for the construction of characters can thus directly contribute to the success of a story. Prior psychology, philosophy, and literature research analysis presents character construction as two stages: **conceptualization**, where the creator develops understanding and ideas about their character's attributes (e.g., physical appearance, personality traits, social status, etc.), and **exposition**, where the creator expresses these attributes through a medium (e.g., through writing) [131].

One form of character construction support is conversation with a chatbot embodying the character [5, 116, 127]. Conversing with a character to develop understanding about them could support conceptualization and inspire exposition [5, 116]. Such an approach can be more engaging than both some traditional and more modern ones (e.g., reading texts [20, 100], filling templates [116], character information generation through computational methods [22, 126, 128], and information visualization through images and charts [81, 123, 136]). Moreover, a similar chatbot can support different levels of control over creative processes, from inspiration only to co-writing

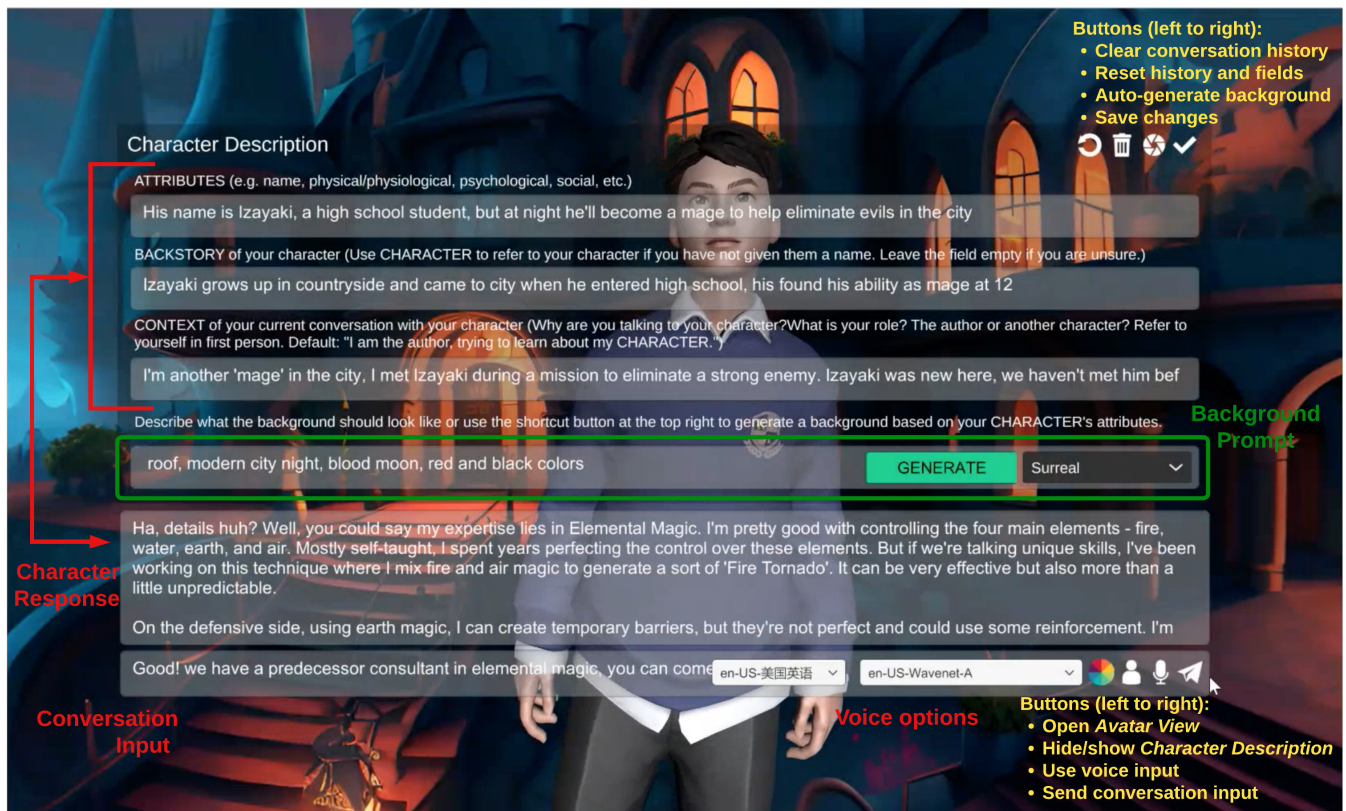


Figure 1: CharacterMeet is a chatbot avatar system aimed at inspiring the entire character construction process (ideating and writing about a character) by allowing writers to converse with their characters. The figure shows the *Conversation View*, which the user opens the software to. The user can converse with their character through voice or text for character and plot inspiration. Anytime, the user can update the character’s attributes (“ATTRIBUTES” and “BACKSTORY”) for them to be reflected in the character’s responses. They can also update the conversation “CONTEXT” for the responses to reflect specific situations and/or to roleplay. The user can visualize their character by customizing the avatar voice or appearance (*Avatar View* in Figure 4) and the context by generating the background through textual description and art style options. Figures 5 and 6 show the relationships between the user and main features.

[5, 116], becoming more widely accepted as an artificial intelligence (AI) tool by respecting writers’ varying senses of ownership [5, 34, 40, 86, 91, 94, 139]. In fact, studies on using large language model (LLM) creative support systems that allow various approaches for storytelling have observed creators adopting approaches where they converse with the systems about their characters or roleplay with them [99, 120, 139]. We focus on conversation for character construction for creative writing specifically since it could lead to deeper insights, which could also augment more general systems. As definitions vary, we define **creative writing** as the composition of original works at least partially through text, which can include the works in Table 3 [55, 79, 80, 88, 101].

Only one research work [116] has studied an implementation of a chatbot embodying story characters specifically for character construction for creative writing. [116]’s chatbot prototype realizes their concept of “*Progressive Manifestation*”, where the writer gradually turns the chatbot into their character through conversation by iteratively updating character attributes for the chatbot

responses to better reflect their imagined character. Such a chatbot that supports progressive manifestation can be more intuitive as it aligns with writers’ own progressive character idea conception process [131].

We build on [116] with several improvements. Firstly, the character construction process could be iterative, like other creative processes [25, 30, 31, 57, 74, 115]. [116]’s experimental design focuses on ideation (conceptualization) only, which does not cover usage patterns that might arise from iterations of conceptualization and exposition. Our experimental design fills this gap by introducing a writing task (Section 4.3), which makes exposition observable. Secondly, advances in LLMs could lead to different user experience findings on character conversation. LLMs are distinguished from other language models (e.g., that of [116]) by their training data across various domains [106], which could allow them to impersonate diverse characters. Limited text generation capabilities in providing “a realistic conversation consistently” could have limited participants’ envisioned use cases of [116]’s system to an early or

less developed character conceptualization stage. We build a system powered by a 2023 LLM: Generative Pre-trained Transformer 4 (GPT-4) [95], whose selection is explained in Section 3.4. Thirdly, conversation context, which could not be specified [116], could also improve realism (as defined in Section 2.2) [5, 116]. We add such a feature. Fourthly, sources of inspiration for character construction are not limited to text. Audio, visuals, and advances in text-to-image generative AI (e.g., Stable Diffusion [112]) could support personalized visualization of the character and conversation context (Sections 2 and 3.2). While characters can take various forms (e.g., animals), for this first study, we balance technical practicality and user freedom of choice by using only human-looking avatars (for better comparability with [116]’s created characters) that are 3D instead of 2D (for possible greater immersion [62, 67]).

Thus, we investigate: *how might conversation with an LLM-powered 3D chatbot avatar in an AI-generated environment inspire the entire character construction process for creative writing?* To answer this, we designed, implemented, and evaluated CharacterMeet, a chatbot allowing writers to progressively manifest characters through conversation while customizing context, character appearance, voice, and background image. Through an online survey, we first informed our design, discovering a need for personalization. We then conducted a user study with 14 creative writers to **examine thinking and usage patterns** of CharacterMeet through observations and participants’ quantitative (assessment questionnaire) and qualitative (semi-structured interview) feedback. Because our design is focused on **supporting** writers with individualized needs, subjective user feedback is our main evaluation criterion. Though we also suggest potential more objective measures (Section 3.4 and 6.3).

Our findings reveal that, firstly, CharacterMeet can support participants by **facilitating iterative character construction processes that even those with more linear usual approaches adopted**. Specifically, participants alternated between writing and idea exploration through visualization on CharacterMeet. Secondly, character exploration through conversation on CharacterMeet, powered by the latest technologies, **supported entire character construction processes instead of only early stages**. Thirdly, CharacterMeet complemented the thinking patterns of writers who had different exploration needs, visualization preferences, and attitudes toward ownership, demonstrating the **need of AI creative writing support tools to personalize** to writers’ varying approaches.

Thus, our key contributions are: (1) the first user study on conversation with chatbot avatars embodying story characters for supporting the entire character construction process in creative writing along with a writing task created specifically for user studies focused on entire character construction processes, (2) the introduction of CharacterMeet, a chatbot avatar that can support writers’ entire character construction processes by personalizing to their varying thinking and usage patterns, and (3) qualitative and quantitative feedback from creative writers with diverse cultural backgrounds, writing experiences (Section 4.1 for both), and creative processes (Section 5.1) that contributes to both personalized AI creativity support tool design and iterative creative process research.

2 RELATED WORK

2.1 Theoretical Background on Character Construction

To create a tool that supports character construction, we need to understand what is considered ‘good’ character construction. Starting with conceptualization, a captivating character is commonly considered to be a **multidimensional** set of attributes, which are mainly related to physiological/physical (e.g., age and eye color), psychological (e.g., personality traits and ambitions), social (e.g., profession and social status), and story (i.e., both the character’s backstory [108, 116, 119, 131] and their role in the current story [35, 85, 119]) dimensions [11, 16, 27–29, 33, 50, 92, 108, 116, 119, 131]. A feature that participants from [116] found helpful was a recommendation system that suggests one of a limited number of attributes. Taking advantage of LLMs’ broader capabilities, we provide prompt suggestions without restricting the attributes to respect varying definitions (Figure 2).

As rich as a character’s attributes might be, the evaluation of a story would depend on how these are conveyed (exposition or “**characterization**” [38, 48, 131]) in different scenes throughout the story [17, 65, 66, 103, 118, 131]. To address the varying definitions of a ‘good’ story, works have attempted to find standards of measures, some related to characterization [14, 90, 110, 130, 140]. They commonly highlight the importance of **indirect** expression of the character’s attributes through a number of “expositors” [131] that can be summarized as “physical appearance, action, thought, dialogue, setting, and symbol” [130] for specific scenes and realistic character development, character changes throughout the story. Moreover, character construction is closely tied to the character’s relationships with other characters [10, 48, 125, 131]. While the users, acting as themselves, could directly inquire about character development and relationships from the chatbot [116], conversation with the character during a specific story scene and roleplay as another character by **defining a conversation context** could provide additional insights for both conceptualization (e.g., greater freedom of character exploration [115]) and exposition (e.g., dialogue under different contexts/scenes [131]). For a specific context, the mentioned expositors also suggest that sources of inspiration for characterization could be other than text (e.g., character visual for the description of physical appearance, location visual for setting, and voice for the description of timbre and pitch during a dialogue). We explore their applications in Section 2.3.

2.2 Character Exploration Through Chatbots

While studies on LLM-powered systems for storytelling in general have observed writers obtaining inspiration for characters through conversation [99, 120, 139], none has focused on the visualization of characters through sources of inspiration other than text nor on progressively manifesting the characters through conversation by updating their attributes. Several research works have focused on user conversation with chatbots to specifically develop understanding of both fictional [20, 76, 116, 127] and non-fictional [100] characters for writers [116], other creators [127], and the audience [20, 76, 100]. [5] has obtained writers’ feedback on the concept of a “conversation-driven approach”, where the writer can talk with

PROMPT: Based on BACKSTORY, suggest 5 ATTRIBUTES other than the ones you have as a numbered list. For each ATTRIBUTE, explain how it relates to BACKSTORY in the format ATTRIBUTE: explanation.

RESPONSE:

- 1. Stealthy: As a retired assassin, Gina developed the ability to move silently and avoid detection.*
- 2. Regretful: Haunted by her past, Gina carries a sense of regret for her previous actions.*
- 3. Skilled fighter: Her career as an assassin required her to be proficient in various combat techniques.*
- 4. Cautious: Faking her own death and living in anonymity has made Gina extremely careful in her dealings.*
- 5. Adaptable: In order to survive her dangerous past, Gina needed to be highly adaptable to changing situations.*

Figure 2: An example prompt and response (generated by OpenAI GPT-4 [96]) that we showed participants. Here, we let them know that they can request specific formats and information when asking for attribute suggestions. “BACKSTORY” and “ATTRIBUTES” refer to *Character Description* fields (Figure 1).

their characters through roleplay to understand them, but did not evaluate any implementation, which can lead to different feedback.

Generally, conversation can support more flexible, personalized, and engaging exploration, essential for creative processes [115]. Conversation with a character as oneself or through roleplay can lead to greater immersion or place illusion (Section 2.3), with other non-textual content (e.g., audio and visuals) enhancing this. The main challenges to user experience are limited freedom of exploration and the **realism** of the characters - how closely the conversation and potential non-textual sources of inspiration reflect both the character’s attributes (e.g., personality) and the context (e.g., factual information [20, 100] and situation-specific responses [116]) that writers have in mind. As both are dependent on the technical, the latest technologies could provide greater freedom of input and produce more realistic output. Our user interface, specifically the layout and input field choices, is inspired by both the mentioned research works and existing industry tools that could be used for character construction through conversation [1, 12, 53, 54, 70, 95].

2.3 Storytelling Inspiration Beyond Text

While no prior research work has focused on non-textual sources of inspiration for character construction for creative writing specifically, works have focused on a variety of potential sources that could inspire storytelling, such as visuals [2, 44, 45, 81, 137, 141], audio [4, 42, 44, 63, 114], smell [42, 43], and eating [133]. [116]’s preliminary survey with 30 creative writers indicates that creative writers draw inspiration for character construction from music, other stories, real life, and online research. In addition, [20]’s findings suggest that character voice could facilitate character exploration. The existence of several potential sources of inspiration aligns with observations that storytellers’ inspiration comes from real-life experience, an amalgam of multi-sensory stimuli [25, 31, 74, 131]. Perhaps closest to real-life experience is the feeling of “being there”, **place illusion**, which works have managed to create [20, 44]. In particular, [44] found that “mild” place illusion (created through location-related visuals on a computer monitor) could inspire creative writing.

Considering technical limitations our potential users might have, we focus on visuals and audio. The chatbot-related works (Section 2.2) that use these to represent their characters do not study their effects on inspiring character construction and provide no option to customize them to the image of one’s own character through the systems’ interfaces. For visuals, we focus on the visualization of **physical appearance** (Section 2.1) through an animated full-body 3D avatar (Section 3.3.4). We chose animated avatars over photorealistic ones as more photorealism might not necessarily lead to more comfortable user experience [77, 87, 113]. As the setting also contributes to character construction (Section 2.1), our system supports the generation of location-related background images (Section 3.3.3). While [81]’s system includes both types of visuals we include, it does not focus on their effects on character construction and is not a conversation-based system. For audio, we focus on **character voice** as it is directly reflective of the character (Section 2.1) and animated oral communication with a virtual human might be a more plausible interaction method, which could provide additional inspiration through illusion [44, 78, 135]. As an equivalent to [116]’s recommendation system, we provide visualization of the various customization options for the avatar (Figure 4) and include an automatic generation feature for the background image (Section 3.3.3).

3 SYSTEM DESIGN

With better theoretical understanding (Section 2), we designed a survey (Section 3.2) to inform our design choices (Section 3.3) through empirical evidence. To do so, we first defined our target user/participant group (Section 3.1).

3.1 Target User Group

Due to differing views on whether the character is an essential element of creative writing, we specify that our target users are those who write passages with characterization from work types in Table 3 or that the user considers relevant. As academic and professional experience might not necessarily reflect creative writing ability and there seems to be no other agreed standard of measure, we do not categorize our participants as ‘amateur’ or ‘expert’ [40].

3.2 Empirical Background from User Research

We present each section of our online survey (implemented on Qualtrics [105]) along with the results then discuss our findings.

3.2.1 Demographics. We recruited 30 creative writers (15 female, 14 male, and 1 non-binary; 28 aged between 22-35 with average age difference of 3.9; 2 aged between 46-60) through popular social media platforms in different countries. We asked about factors that could affect creative writing habits: cultural background (“Which countries’ cultures have you spent the longest living, working, and/or studying with?”) [56, 89], creative writing education/experience, and written work type(s). The participants were from different **cultural backgrounds** (21 chose *China*, 6 *Canada*, 4 *United Kingdom of Great Britain and Northern Ireland*, 3 *United States*, 1 *Philippines*, 1 *Bulgaria*, 1 *France*, 1 *Germany*, 1 *Kenya*, and 1 *Russian Federation*). For **creative writing education**, 8 chose *At least a bachelor-level degree in literature, creative writing, fiction, or related fields*, 13 *Some classes (but no degree)*, 1 *Some workshops*, and 8 *Informal (e.g., consulting writing guides by yourself)*. For **creative writing experience**, 10 reported full-time professional experience and 15 part-time. For **work type(s)**, 25 participants chose T1, 9 T2, 9 T3, and 8 T4 (Table 3).

3.2.2 Sources of Inspiration. We ask a question (“From where do you usually draw inspiration to design and write about your characters?”) where the participant can choose multiple options we designed based on Section 2. We gave a specifiable “Other” option. All chosen options are shown in Figure 3. **In addition, among the option(s) they have chosen, each participant has chosen an option that is not “written/textual characterization in another narrative”.**

3.2.3 Ownership. We also investigated different writers’ attitudes toward ownership of their work when using a generative AI creativity support tool (“Assuming generative AI can produce satisfying results for any use case, how much would you let it contribute to your creative writing work for you to still feel like you are the author (assuming that you do not have to worry about copyright or plagiarism issues)?”). The participant completes the sentence “I am fine with generative AI generating:” with one of the choice options we provide. 3 participants chose *the entire work based on little information I provide it (e.g. a title and a brief summary)*, 5 *the entire work based on a detailed outline I provide it*, 4 *large portions of the work but not the entire work*, 7 *only small parts of the work I have difficulty writing*, 4 *corrections and improvements for a work I need reviewing*, 5 *content that I use as inspiration only*, and 2 *nothing. I do not want such a technology to contribute to my work*.

3.2.4 Discussion. Participants were from diverse backgrounds (Section 3.2.1). For sources of inspiration, all are inspired by at least one of: conversations, character visuals, location, or voices. This suggests that related system features (Section 2) might serve as inspiration to some writers. In line with past research [40, 86, 139], participants have different attitudes toward ownership and individualized needs, highlighting the need for a design that provides flexible support [122].

3.3 User Interface and Features Design

Based on prior research and our survey, we derived one major goal for the design of a chatbot system aimed at supporting the entire character construction process through conversation: personalizing to writers’ varying character construction approaches in terms of (1) sources of inspiration, (2) writing processes (e.g., order of stages and attitudes toward ownership), and (3) written content (e.g., genres, work types, character attributes and exposition styles, etc.). To facilitate personalization, we designed the main features of CharacterMeet to provide as much freedom of input as possible (i.e., freeform text input for generative AI features and various options for other customization features) while minimizing our influence on the usage of different features (e.g., by not enforcing any particular order of use and providing prompt suggestions touching various use cases to facilitate discovery while not enforcing any specific format or content restriction).

3.3.1 Interface Overview. CharacterMeet has 2 views: the *Conversation View* (Figure 1), which the user opens the software to, and the *Avatar View* (Figure 4), which displays the avatar customization feature for visualization of the physical appearance of the character. Relationships between the user and main features are shown in Figures 5 and 6.

3.3.2 Progressive Manifestation Through Conversation. The *Character Description* window supports the progressive manifestation of the character through conversation with the following freeform input fields: (1) **ATTRIBUTES** (any attributes of the character, except for backstory, with examples in Table 1), (2) **BACKSTORY** (significant events in the character’s past life), separated from other attributes for visibility as this could be relatively longer, and (3) **CONTEXT** (context of the conversation, including the setting and the user’s role, as oneself or another character). For ease of update, CharacterMeet has save/clear buttons as shown in Figure 1.

3.3.3 Background Generation. CharacterMeet uses text-to-image generative AI for personalized visualization of the setting (background image). CharacterMeet supports background generation in two ways with various art style options to enhance place illusion: (1) self-defined generation, for which the user describes where they envision their character to be, and (2) automatic generation. Created based on the notion that ‘good’ setting exposes the character’s attributes (Section 2.1), this latter feature automatically generates the background based on **ATTRIBUTES**, **BACKSTORY**, and **CONTEXT**. We intend for this to inspire the user by revealing details about the character and through place illusion [44, 135].

3.3.4 Avatar View. As we found no existing suitable text-to-3D generative AI for the avatar and developing one would not be worth the efforts for our research aim, we integrated a manual animated 3D avatar customization platform (introduced in Section 3.5). The *Avatar View* (Figure 4) combines the layout of the customization platform with our own to provide a preview of the avatar with our background.

3.3.5 Voice Input/Output. We include voice options both to directly inspire conceptualization/exposition and create place/plausibility illusion (Section 2.3).

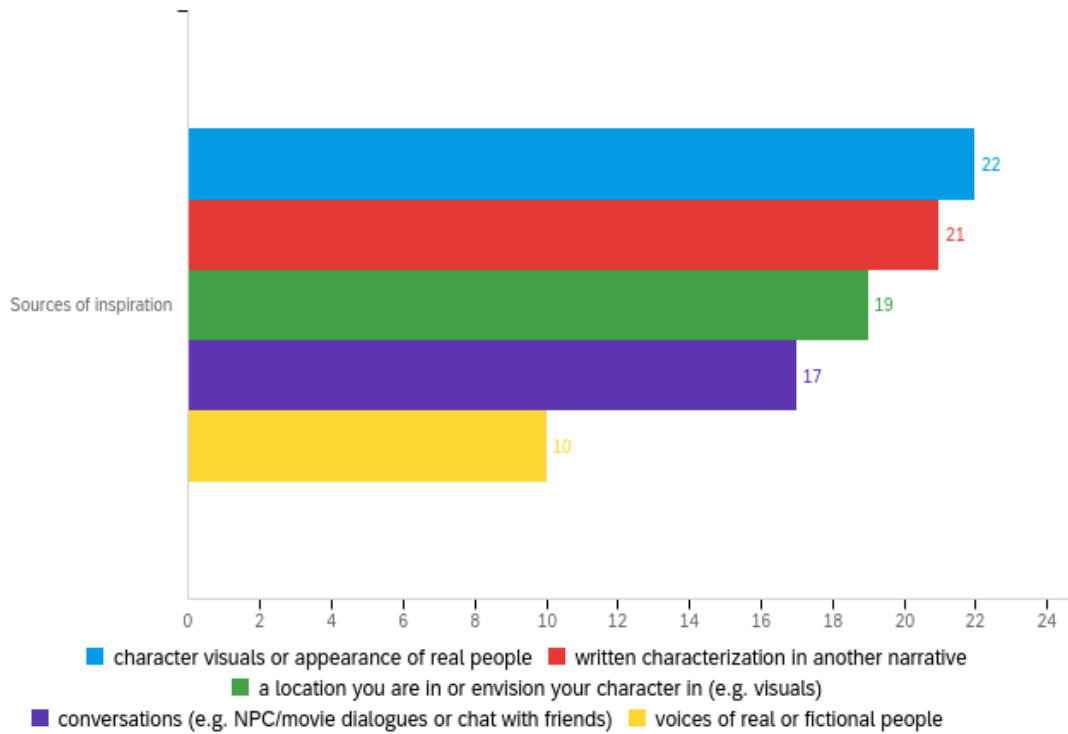


Figure 3: All options selected for the sources of inspiration question (“From where do you usually draw inspiration to design and write about your characters?”) in Section 3.2 in descending order of participant counts (option labels in same order).

3.3.6 Other Commands. CharacterMeet’s text generation capabilities can provide support beyond conversation. Inspired by prior studies on writers’ experience with language models [71, 99, 116, 139], we designed and showed command prompts for writing support (e.g., attribute and context suggestions) to participants (Section 4.2.1).

3.4 Model Selection

Focused on personalized support, our LLM-powered system mainly has to generate responses impersonating diverse characters while providing miscellaneous creative writing support. Given writers’ individualized (possibly contradictory [122]) support quality criteria, we choose GPT-4 based on its greater potential in satisfying varied needs by analyzing a breadth of capabilities related to domain knowledge (for diverse character backgrounds), impersonation, and language performance assessed through different methods. Specifically, GPT-4 performs better than or similarly to other LLMs for benchmarks on language, reasoning, and domain expertise (e.g., multilingual Massive Multitask Language Understanding on 57 academic/professional subjects, HellaSwag, and WinoGrande, both on commonsense reasoning [96]; dialogue generation benchmark [73]; manual analysis of responses covering reasoning, factual knowledge accuracy, and writing tasks [7]), and for evaluation on creative writing and impersonation (e.g., plot logic, originality, character exposition, and structural elements, such as grammar and formatting [41]; response consistency with human expectations of personas

[52, 58, 134]). GPT-4 also performs similarly to humans in language, domain knowledge, and creative tasks (e.g., SAT and AP tests [96]; Torrance Tests of Creative Thinking [46]; creative writing, assessed through mimicking famous authors’ styles [9, 37] or through comparison with works written by humans [41]), suggesting potential in providing support on par with humans.

To be more confident in GPT-4’s capabilities, we also invited 5 creative writers (one without LLM usage experience) to remotely experience, improve, and discuss our prompting setup (Figure 6), designed by one of our authors who has creative writing, psychology, and LLM use experience. Writers filled the attribute input fields based on characters they might write about then conversed with and adjusted attributes until generated responses “feel like” what their characters would say without and with a context. Although they could give up, all writers (the one without LLM experience with prompting advice from us) ended with character responses that are **satisfactory** - that reportedly could be integrated into their stories with minor coherence modifications (not accounting for ownership, copyright, or factual accuracy concerns). Characters created belonged to different genres, namely modern realistic fiction, fantasy, historical fiction, and science fiction. This suggests that our prompting setup powered by GPT-4 could reasonably satisfy individualized needs.

As finding a commonly agreed upon output quality measure for a support tool that personalizes to individualized creative needs

can be challenging [122], the writers and our aforementioned author with creative writing experience discussed possible measures, agreeing that the **number of attribute adjustments (sessions between conversations) before a writer obtains satisfactory responses** (3-6 for us) could be an objective measure of efficiency and output quality, whose minimization reflects greater user satisfaction. As we found no comparable creative support tool research, our data could serve as reference for future works.

Our discussion also revealed major possible breakdowns: (1) **character world knowledge inaccuracy** (real or fictional) and (2) **content filtering**. For (1), LLMs could hallucinate inaccurate information [51, 95] (e.g., invent a plausible but false historical event for a historical character). While fine-tuning could mitigate these issues, for a first study, we favor technical practicality. For (2), LLMs might refuse to generate responses that could be considered sensitive [7] (e.g. for a character expected to talk offensively). To respect usage policies [98] and our researchers, we took no additional measure. We also noticed increasing response time (from 3-4 to 9-10 seconds on average) but imposed no additional token limit (default: 8k) to accommodate varying conversation lengths, which writers found more important. We cover future design implications in Section 6.2.

3.5 Implementation Details

We developed CharacterMeet as a desktop software in the Unity game engine. For the conversation, we used OpenAI GPT-4 API [96] (Figure 6). For the background generation, we used Blockade Labs' Skybox AI API [69], which provides access to a skybox generator. The generator is powered by Stable Diffusion, which has been shown to surpass other image generation models in terms of image generation accuracy and resolution [112]. We chose [69]'s platform mainly due to its diversity of art styles, which could be suitable for different story moods and genres. For the auto-generation feature, we send a prompt to GPT-4 for a response prompt suitable for the skybox API's input field. For the avatar customization, we integrated Ready Player Me's Avatar Creator interface into our software through a provided Unity SDK [82]. We chose Ready Player Me due to its available avatar animations and its variety of clothing, face, and body appearance options, which could support the visualization of characters from various cultures, eras, and story genres. For the voice input and output, for potential participant cultural diversity and character diversity, we chose a private speech synthesis service with several languages and several voices for each language. The service API converts GPT-4's text output into voice output and transcribes the user's voice input into text input for GPT-4.

4 USER STUDY

We conducted a user study with 14 creative writers to discover thinking and usage patterns through quantitative and qualitative feedback on our system. We first pilot-tested our study design with 3 creative writers, who were unaware of the research and had varying degrees of generative AI experience (including none).

4.1 Participants and Apparatus

We recruited 14 participants (anonymized as P1, P2, P3, etc.) through online advertisements on popular social media platforms, university communities, and word-of-mouth in various countries (China and Western countries). In addition to general demographics information - gender (8 female and 6 male), age (23 to 35 years old with an average difference of 2.7 years), and level of education (6 with "University bachelor's degree" and 8 with "Graduate or professional degree") - we collected other information that could affect user feedback. Firstly, participants reported their **cultural backgrounds (same question as Section 3.2)** since character construction is tied to one's exposure to cultures [49, 125, 131, 138], participants' assessment of character realism could be influenced by cultural biases generative AI might have [18, 93], and cultural experience could affect usage patterns of technologies [72]. Secondly, participants reported their **creative writing education and experience (same as Section 3.2)** since familiarity with the writing process might affect the support type sought from CharacterMeet. Thirdly, participants reported their usual **written work type(s) (same as Section 3.2)**. While the writing task is purely text-based fiction, a participant's past focus might lead to different usage patterns. For instance, those who focus on nonfiction might require greater factual accuracy, on multimedia works, more variety for visuals and audio since their works might include these, and on interactive works (commonly characterized by choices leading to different story branches [6, 13]), more time to explore alternate scenes. Fourthly, participants reported their **technological proficiency** (text-based and image-based generative AI experience), which could affect expectations, especially since prompting for a generative AI is a trial-and-error process [132].

For the **cultural background**, 12 participants chose *China*, 2 *Canada*, 1 *United States*, 1 *Philippines*, and 1 *United Kingdom of Great Britain and Northern Ireland*. Like for Section 3.2, participants are from diverse cultural backgrounds with many having experience with non-Western cultures. This could be especially interesting given the potential biases generative AI might have toward groups considered to be minorities in Western societies [93]. For **creative writing education**, 5 participants chose *At least a bachelor-level degree in literature, creative writing, fiction, or related fields*, 6 *Some classes (but no degree)*, and 3 *Informal (e.g., consulting writing guides by yourself)*. For **creative writing experience**, 5 participants have more than 3 years of full-time professional creative writing experience, 4 had part-time experience, and the rest have only written as a hobby. For the writing of story passages with characterization specifically (see Section 3.1 for the definition of our target users/participants), 2 participants had less than 1 year of experience, 3 1-3 years, 7 3-6 years, and 2 over 10 years.

All participants are fluent in English but have different preferred languages for creative writing. While our aim might be irrelevant to the language used, the English performance of GPT-4 and of our image generator used might outperform the performance of other languages ([95] for GPT-4 and manual tests for the image generator). We balanced possible language barrier biases and the technical biases of generative AI by letting participants use their preferred language and switch to English for CharacterMeet if they

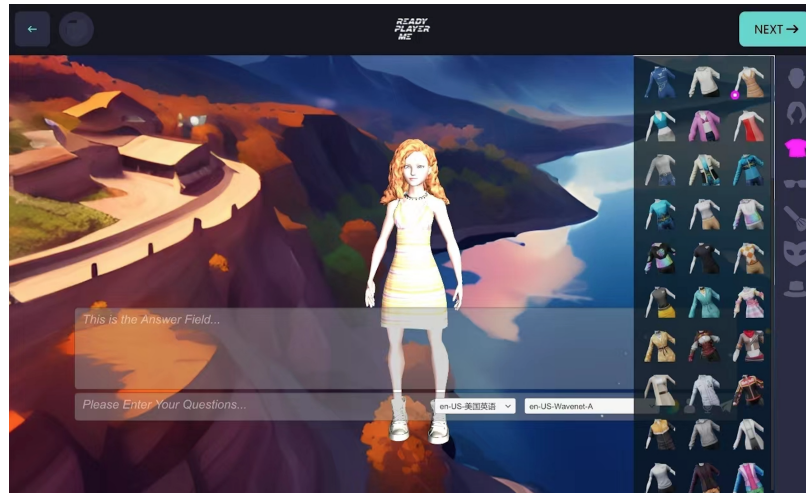


Figure 4: Avatar View of CharacterMeet that integrates a third party platform (Section 3.5). The user can manually customize the avatar’s face, body, and clothing appearance while visualizing all the customization options (as seen on the right).

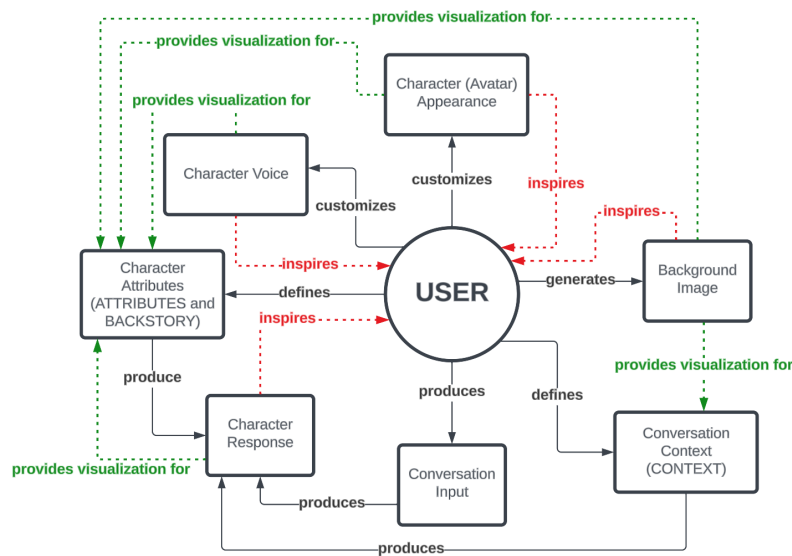


Figure 5: Illustration of relationships between the user and main features of CharacterMeet. The user can visualize character appearance, other attributes, and conversation context through customization of an avatar, a generative AI background, and voice output options.

experienced inaccuracies. **Participants who did not initially use English only changed the language for the image generator.**

Table 2 shows each participant’s language(s) used (“Language(s)”), the types of works they most often write (“Type(s)”), and the type of generative AI they have used for creative writing if applicable (“Generative AI”).

4.2 Procedure

We intended for each participant to participate in a single experiment session as follows.

4.2.1 Introduction. One researcher conducted all sessions in-person or remotely through video calls depending on the participants’ situations. Upon the participant’s signing of the consent form, the researcher introduced our research goals and the different features of CharacterMeet. The researcher also introduced prompting techniques and sample prompts (example in Figure 2) to the participant to diminish learning costs based on needs we identified during the pilot study. The participant could follow along with the system. For in-person sessions, the participants used software already installed

Format	Example
full sentences only	CHARACTER's name is Aria. She is 22. She is a college student who is cold and distant. She does not have many friends. She is very loyal to her friends.
keywords only	name: Aria; age: 22; college student; cold and distant; does not have many friends; loyal to friends
mixture	name: Aria; age: 22; college student; She is cold and distant, but she is very loyal to her friends.; doesn't have many friends

Table 1: Examples for different possible formats of the ATTRIBUTES field input (“e.g., name, physical/physiological, psychological, social, etc.”). “CHARACTER” refers to the defined character.

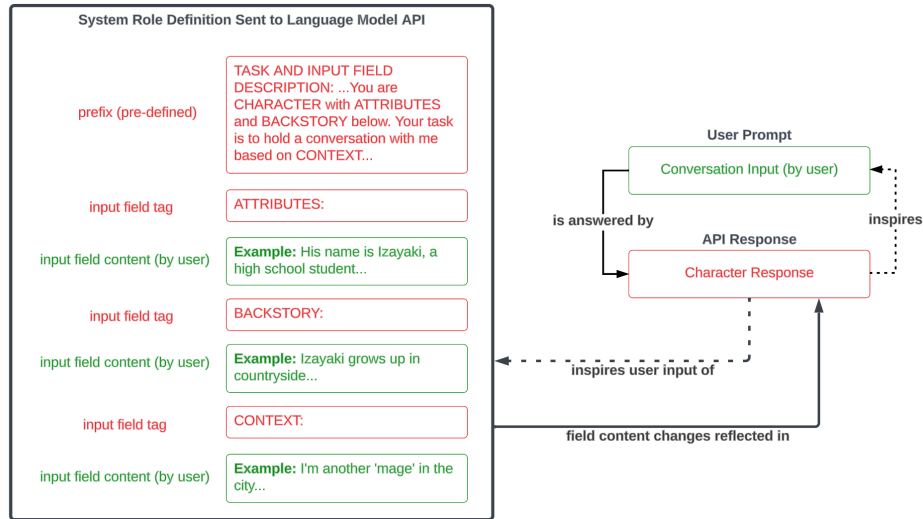


Figure 6: Illustration of the prompting setup for CharacterMeet with tags corresponding to input fields in *Character Description* and example input content from Figure 1. The language model’s API request body [97] supports defining the system (concatenation of prefix, tags, and input content) separately from the conversation body. Whenever the user sends their conversation input, it is sent to the language model along with the saved changes of the system definition. The model then generates a response based on all this.

on our computer. For remote sessions, we sent CharacterMeet to the participants.

4.2.2 Warm-Up. The participant explored different features for character building through CharacterMeet and asked questions if needed.

4.2.3 Writing Task. Once the participant explored CharacterMeet to their satisfaction, the researcher introduced the writing task (Section 4.3), which the participant was encouraged to complete with the support of CharacterMeet. The researcher suggested that the participant can think aloud but minimized interruptions during the task. The participant could request assistance from the researcher at any time. Since participants might have different writing habits and needs, we imposed no time or word limit to observe a writing process that is as reflective as possible of the participant’s usual writing process. Participants all wrote in Microsoft Word, which was a word processor they were all familiar with.

4.2.4 Post-Task. The participant evaluated CharacterMeet through an online questionnaire and a semi-structured interview. The interview focused on their usage patterns (“*Did your interaction with CharacterMeet inspire any part of your writing? How?*”), possible differences compared to their usual writing processes (“*How similar is writing with the support of CharacterMeet compared to your usual writing process?*”), character realism (“*Did interaction with your character feel close to what you imagined your character to be like? How did this affect your writing process?*”), further use (“*Would you include CharacterMeet in your writing process? How?*”) and “*Is there any feature you would like CharacterMeet to have?*”), and ownership (“*How do you feel about your ownership over what you have created today?*”).

We conducted all sessions within a week in August 2023. Because we imposed no time limit, the study is of varying lengths for different participants. All but 2 (P1 and P9) completed it in a single session of about 1 hour and a half to 4 hours. P1 completed the study in two sessions, one of about 2 hours (until the end of the writing task) and another of about 1 hour. P9 completed the study

ID	Language(s)	Type(s)	Generative AI
P1	CN	T1, T4	text, image
P2	CN	T1, T4	text, image
P3	CN	T1, T2	text
P4	EN	T1	none
P5	EN	T1	text
P6	EN (image)/CN	T1, T3	image
P7	EN	T1, T3	text, image
P8	EN (image)/CN	T1, T4	none
P9	EN	T1	none
P10	EN (image)/CN	T1, T2	text, image
P11	EN	T1, T2	text, image
P12	EN (image)/CN	T1, T3	none
P13	EN (image)/CN	T1, T2, T3	text, image
P14	EN	T1, T2	none

Table 2: Participant information. “CN” means “Chinese”. “EN” means “English”. “(image)” indicates the final language the participant used for image generation. Table 3 shows the corresponding choice options for the abbreviations used for “Type(s)”.

in a session of about 3 hours (until the end of the writing task) and another of about 1 hour. Participants spent from about 30 minutes to 2 hours on the writing task. Our study has received ethical clearance from our institution. We also provided a compensation of about 27 USD to each participant upon completion.

4.3 Writing Task Choice

To observe the potential of different system features, we aim for the participant to experience expositors as comprehensively as possible (Section 2.1) through the writing task. As we found no suitable research writing task for our intended study time length (one session), we designed our own task by balancing task complexity, task relevance to the writer’s usual writing, and possible biases due to the writers’ differing backgrounds. For task complexity, we had to choose a task that was short enough for one session but focused enough for the participant to experience the entire character construction process and touch different expositors [140]. For the former, we asked the participant to **write about a character they had never seen before or written about**. For the latter, based on character writing exercises and advice [11, 47, 61, 68, 84, 92, 131], we designed our task as follows: the participant had to **write a story scene focused on the depiction of one main character who learns about bad news**. Writing about a character’s reaction to bad news has often been used as an exercise to practice psychological depiction. Because conflict involving the main character (e.g., bad news) is an essential part of a narrative [109], the task is relevant to our participants’ usual writing processes, potentially leading to insights on how they would integrate CharacterMeet outside of the experimental setting. To reduce biases due to unfamiliarity, we imposed no additional story setting requirement.

4.4 Measures

We obtained our data mainly from assessment questionnaires, transcriptions of the semi-structured interviews, and observations with

think-aloud user feedback. For assessment questionnaires, for comparability with [116], we let the participants fill out *Creativity Support Index* (CSI) questionnaires. For transcriptions, we transcribed all user feedback. Video recordings of the sessions were reviewed for possible observations that were not included. Post-interview follow-up questions were asked when needed. For the interviews and additional content, 11 participants communicated in Chinese (Mandarin) while 3, in English. We manually translated the Chinese content into English. Two researchers then coded the data separately to extract themes [23] and met to achieve consensus.

5 FINDINGS

Using thematic analysis, an approach for identifying, analyzing, and reporting patterns within qualitative data [23], we analyzed all transcripts and identified four key themes. We present them (Sections 5.1, 5.2, 5.3, and 5.4) then the CSI results (Section 5.5) along with representative participant quotes (Participant ID in brackets) and relevant observations.

5.1 Iterative Character Construction Processes

We found that CharacterMeet’s features facilitated both iterative usage patterns between writing and using an AI and iterative character construction processes - iterations between writing (exposition) and exploring ideas through visualization on CharacterMeet (conceptualization) - for all participants, including those who reported having more linear usual processes (P6, P7, P8, P10, P13, P14). Specifically, we identify two features reported by most to facilitate iterative character construction processes: the *Character Description* view, which allowed easier character attribute and context (both description and visual) updating, and the definable conversation context, which motivated participants with more linear usual approaches to adopt more iterative ones. For instance, P8, who usually wrote only after planning their entire plot, found using an iterative process with CharacterMeet to be “more productive”, because they can explore possibilities at different points of their story.

5.2 Exploration Across Entire Character Construction Processes

CharacterMeet supported personalized use cases for idea exploration across both less and more developed character construction stages, contributing even to exposition. By exploring their characters’ reactions through visualization of various conversation situations, participants mentioned obtaining inspiration (1) for new attributes (e.g., details of the backstory, name, personality traits, cultural background, etc.) during both initial brainstorming and writing, (2) for dialogue, character description, and relationships during writing, and (3) for the plot.

For instance, P6 (Figure 1) obtained inspiration for their character’s backstory (the magic society). P12 (Figure 7d) obtained inspiration for both the character’s nickname (conceptualization) and their description based on their nickname (exposition). In Figures 7b and 7c, participants explored the reactions of characters through conversation, which directly contributed to their dialogue writing (exposition). For relationships (mainly conceptualization), P4 explored their character’s relationship with their mother by comparing character responses when P4 acted as an author versus when

Abbreviation	Corresponding Choice Option
T1	short stories, novels, and other forms of fictional non-interactive, mainly text-based works
T2	mainly text-based narrative nonfiction (e.g., memoirs and biographies)
T3	interactive fiction, visual novels, game scripts, and other forms of texts for usually interactive works
T4	scripts and texts for non-interactive multimedia works (e.g., movies, animation, and TV series)

Table 3: Abbreviations used for “Type(s)” in Table 2. The choice options are for the question “What types of works do you most often write passages with characterization for?”. In addition, we included an “other” option that participants can specify.



(a) Example taken from the recorded writing task of a participant and showing CharacterMeet’s interface. The participant mentioned that “what really helped [them] to have an immersive experience is actually the background”. Because it has “the same vibe” as the “coffee shop” the participant imagined, “especially when the avatar is speaking to [them]”, they feel “as if [they’re] actually there”. While they’re “not drawing anything specifically from the background”, they feel “more inspired”.

Figure 7: Examples taken from our recorded experiment sessions.

they acted as the mother. P3 and P6 learned about their characters’ attitudes toward other characters through roleplay while P9 and P12, by conversing as themselves. Participants have also reported obtaining inspiration for the plot by conversing with their characters. For instance, P9 asked their character “what the bad news was”.

5.3 Varying Visualization Needs for Visuals and Voice

Visuals and voice mainly inspired through immersion or a sense of ‘being there’ (13 participants, not including P7; example in Figure 7a) instead of being directly used for character description (P5, P7). Participants expressed expectations for CharacterMeet to provide visualization at degrees of vividness that can complement their idea conception without limiting them. Such degrees varied among participants for different features, reflecting diverse customization needs shaped by diverse writing experiences and cultural backgrounds (Section 4.1).

5.3.1 Art Style. Preferences for the avatar and the background styles varied in terms of photorealism, with most preferring a more “cartoonish” style because a more photorealistic style could limit their imagination (Figure 7f). The rest preferred a more photorealistic style, which is more reflective of the character appearance they envision for their usual work types (Figure 7e). Most participants

preferred coherence between the character and background styles (Figure 7e) while others did not mind or preferred otherwise. P9, in particular, preferred “a cartoonish avatar and a photorealistic background”. They explained, “The background being more photorealistic is important for you to immerse yourself into the story environment. On the other hand, how photorealistic, how detailed, the character looks is not that important. The character is more abstract.” Most participants mentioned that art style choices depend on genres and/or work types.

5.3.2 Voice. Most participants wanted the voice to sound “less AI” as it can be “distracting” (P9) while the rest found voices that represent their characters enough to enhance immersion. Some participants (P2, P4, P6, P11) also mentioned specific requirements for accents, timbre, and pitch diversity to help visualize characters from diverse backgrounds.

5.3.3 Vagueness of Visuals. Participants mentioned different preferences for the vagueness of the avatar’s form and the visibility of the background. For the avatar, while some wanted more customizable features for a more detailed depiction of their character, others preferred vaguer representations than our current one. Some explained that their preferences were related to their usual works. For instance, P12, a romance novelist, mentioned that they usually intend for their audience to focus on “the feel” of their

After working out a whole day in the gym, he finally decided that he would be the one who spoke out and be responsible for all the consequences, if necessary.

“You know what? Just ‘cause these wet blankets I call roommates don’t want to grow a spine, doesn’t mean I’ll stay quiet. I’m sick of beating around the bush. If these fools at the top think they can suppress our basic rights, they’ve got another thing coming. Apparently the only language they understand is ‘disruption’, well, then I’m gonna be their professor. This PhD student is about to give them a primer on student rights. Fuelled by pure resentment and a gallon of protein shakes. Brace yourselves!”

It’s rather a difficult choice because of the state’s law of obedience. It was issued at the dawn of the AI era. At first, the school just prohibited student’s daily access to ChatGPT. But soon, an embodied AI machine was invented and plugged into every student’s personal laptop. It was not long before such a device became somewhat an all-day surveillance: some cases are reported that students who look up irrelevant contents to their research during work hours are under arrest and disappear forever.

“See, this is the sort of Orwellian nightmare I am talking about. It’s terrifying enough to have AI systems in our devices, but now they’re even turning into our bloody wardens. Watching us, scanning our every action, ready to throw us into the pits if we dare to deviate. All enforced under the guise of state’s law of obedience, the technological equivalent of a damn dictator. You know something’s messed up when the tools for empowerment become the chains of subjugation. Makes you wonder what we’re doing here, with these PhDs or, hell, even living in this tech-dictated dystopia. If any of those disappeared students are listening, just know that this fight isn’t over. The muscle-head is on your side.”

Jason tried to say something but most of the contents disappeared the minute he typed them in his mobile device.

(b) Extract of a participant’s writing task response. The participant took turns writing part of the scene then asking CharacterMeet to generate a dialogue that followed. The participant explained that they wanted their character to “have a say” in their story. The participant mentioned that they also based part of the plot on the dialogue generated at each iteration. The highlighted portions are responses taken from CharacterMeet (generated by GPT-4).

There I realized I just finished my coffee and needed a refill. I stood up and went up to the counter to order a second, only to be standing right behind her.

“That’s a complicated drink, isn’t it?” I asked, and she replied, “Haha, well, isn’t life complicated enough? I guess I just enjoy a bit of extra flavor in my cup. It’s one of the small joys that make my mundane corporate days a little bit brighter. Do you always order the simple black coffee?”

She caught me by surprise.

(c) Extract of a participant’s writing task response. The highlighted portion is a response taken from CharacterMeet. The participant, who is not female, explained, “If I have to design the dialogue between the guy and the woman, the guy side, I think I’ll be just fine, but from a woman’s perspective, I probably don’t know how they will respond or the tone that they use or how they will take a certain type of questions.”

“小狐狸……”猎人的声音从右边的密林里传来。

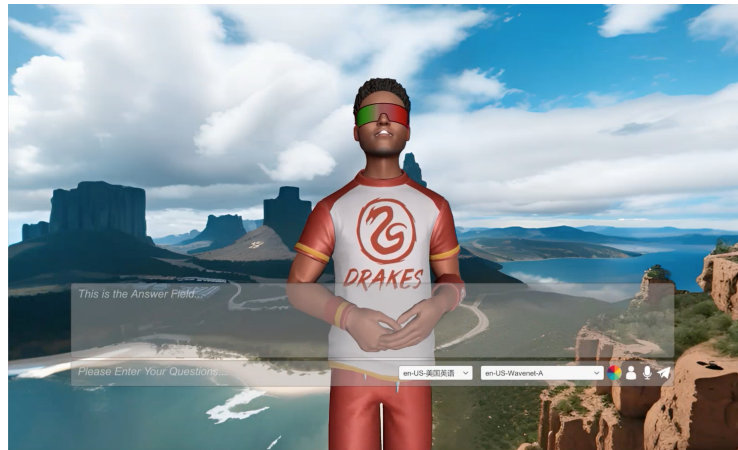
几个身穿警服的人押着猎人出现在亚历克斯的面前，亚历克斯瞪着眼睛，圆圆的眼睛像森林里的狐狸一样大而且亮，这正是猎人喜欢的东西——狐狸的眼睛。

(d) Extract of a participant’s writing task response (in Chinese), where someone calls the main character by their nickname “Little Fox” and a description comparing the character to a fox follows. The participant mentioned that this description was inspired by their conversation with their character in CharacterMeet, where the character suggested “Little Fox” as his nickname and explained that it was “because foxes have large, bright eyes”.

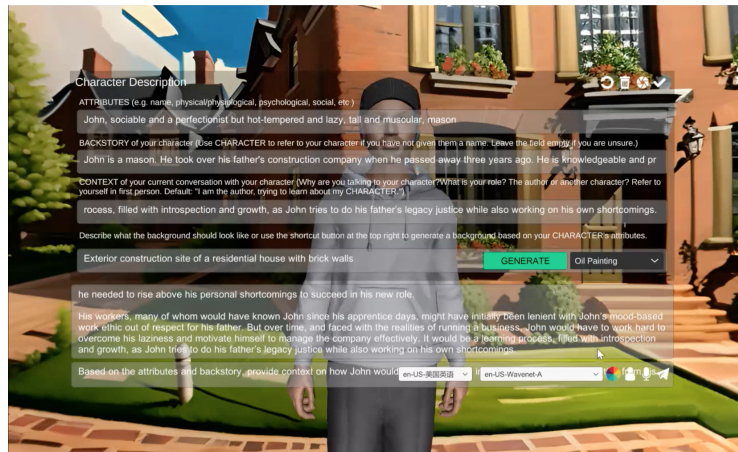
Figure 7: Examples taken from our recorded experiment sessions (cont.)

characters, preferring a vague character representation with “only the distinctive features being clear”. On the other hand, P2, a live-action screenwriter, found realistic character appearance details more crucial to their casting. For the background, P4 found that the background enhanced immersion during the brainstorming stage but preferred no background during writing as it “limit[ed] your imagination”.

5.3.4 *Other Demands*. Participants expressed additional needs for more dynamic visualization, which include having multiple avatars conversing simultaneously and backgrounds that automatically change based on the conversation.



(e) Example taken from the recorded writing task of a participant showing CharacterMeet’s interface. The participant mentioned that they preferred a more photorealistic style (like the background) than the other “cartoonish” styles for both the avatar and the background. The participant explained that they usually “work with many videos and photos”. The participant also emphasized coherence between the background and avatar styles, “If we use a realistic style for the character, the background would be better with a realistic style.”



(f) Example taken from the recorded writing task of a participant and showing CharacterMeet’s interface. In contrast to the participant from Figure 7e, this participant preferred a non-photorealistic style (the background image and the avatar as they are). This participant explained, “I like the current style, because if [the visuals are] too detailed, your imagination might be limited.”

Figure 7: Examples taken from our recorded experiment sessions (cont.)

5.4 Complementary Capabilities

Focused on personalization, CharacterMeet’s breadth of capabilities were perceived as complementary to individual creative processes from the following angles.

5.4.1 Ownership. All participants found use cases for the AI features that align with their senses of ownership. Specifically, participants ranged from willing to directly use content generated by CharacterMeet (even to “make money” from published works, quoting P14) if they do not have to worry about copyrighted content generation to only wanting to use it as inspiration (because it would feel like “cheating” to not “actually write” by themselves, quoting P4). Most were in between, considering rephrasing generated content into their “own writing style” (quoting P5).

5.4.2 Complementary Use Cases for Various Work Types. While our study is focused on T1, participants envisioned use cases for different work types (Table 3). For creative writing in general, participants reported that CharacterMeet could augment their already existing processes and/or increase productivity mainly by saving their usual research time and helping them during writer’s blocks. As P4 mentioned, “[A]ssuming that the information it gets is from actual evidence, the research part becomes easier because you can just use this tool.” For greater factual accuracy, P11 also suggested adding a knowledge bank customizable to domains and language nuances across cultures (e.g., speaking habits), which could help with simulating interviews for their non-fiction works (T2). P7 envisioned using CharacterMeet to explore different story branches

resulting from the main character’s choices in multi-ending story-heavy games (T3). P1 and P2 would use CharacterMeet to visualize characters for screenplay scenes (T4). P5 also found the combination of customizable art style and character interaction inspiring for visual storytelling.

5.4.3 Intent. CharacterMeet’s complementary capabilities have been mentioned to distinguish it from tools that support co-writing only by aligning better with writers’ intent. As P6 recalled, they had only seen others “generate entire stories based on keywords, which do not follow [P6’s] thoughts” while CharacterMeet’s ability to “provide inspiration based on what [P6 has] already created or add details based on it is a feature that [they had] wanted”.

For more comprehensive insights, we analyze the extent to which each participant compromised their intent. Given individualized task response lengths and edit types, we focus on the more classifiable content integration approaches of participants, which were more reactive (i.e., CharacterMeet content steering participants’ conceptualization/exposition in different directions than they intended) or proactive (i.e., CharacterMeet only used to enrich clear ideas that participants had). For conceptualization, all participants mainly modified attributes to make the character talk, look, and sound more like a character they had in mind (proactive), with some changes based on conversations (some reactive decisions). For exposition, half let CharacterMeet content shape their character description and/or plot to varying extents, with P1, P7 (Figure 7b), and P8 adopting such reactive approaches throughout their entire exposition. The rest only used CharacterMeet to add details (proactive). Participants who made reactive decisions mentioned that they were “surprised by” (quoting P5) and trusted CharacterMeet’s ability to portray their characters better than themselves.

5.5 Creativity Support Index

Our prototype achieved an average of 80.02 out of 100 (standard deviation of 9.29) for the overall CSI score while [116]’s final prototype scored 76.96. Due to data availability for [116], we run no statistical test. Instead, we compare the different factors of the CSI, which is the suggested method for comparing two different creativity support tools [19]. Our comparison is shown in Table 4. Like for [116], participants also found “Exploration” to be the most important factor, followed by “Expressiveness” and “Immersion”. CharacterMeet scored higher for the last two but lower for “Exploration” for the average factor score. We note that our sample size is about double that of [116] and possibly more linguistically diverse. Our user study task also involves writing. All could have led to more diverse perspectives, influencing both factor scores and counts. Though consistency still suggests possible trends (implications in Sections 6.2.2 and 6.3).

6 DISCUSSION

6.1 Support for Entire Iterative Creative Writing Processes

With broader capabilities [106], LLMs could surpass other language models in personalizing to creative writers’ individualized needs. Studies on usage patterns of other implemented LLM-powered writing support systems [21, 71, 86, 99, 120, 122, 132, 139] have observed

the necessity of supporting iterations between what the writers are working on (e.g., outline or main text) and idea exploration with the tool (e.g., obtaining suggestions). We add to the literature with three main novelties.

Firstly, we demonstrate that **an LLM creative writing support tool focused on progressively manifesting writers’ characters studied through entire writing processes (instead of specific stages)** could also lead to iterative usage patterns that correspond to iterative thinking patterns (Section 5.1). Given that iterations between idea conceptualization and exposition have been observed during creators/writers’ entire usual creative processes [30, 31, 74, 115] - with [30] rejecting that writing is purely linear - **our findings provide both study and system design insights for LLM-powered creative writing support tools that complement writers’ entire varying usual writing processes** (Section 6.2). Specifically, personalizable support could increase AI acceptance by respecting varying senses of ownership, writers’ intent, and content diversity/quality expectations, but practical applications might be limited by copyright concerns (Section 5.4.1). Research could focus on plagiarism likelihood or on prevention algorithms (e.g. what [126] claims to have). Despite recent lawsuits for AI training on copyrighted content [8], participants did not mention these, possibly focusing more on usability. Participants also did not mention integrating AI content out of convenience (e.g., task/time constraints), a prior observation [24, 122], but believed that intent compromises could improve their character construction (Section 5.4.3). Research could investigate the views of more specific writer and reader groups (e.g., creative writing class instructors and commercial novel readers) on copyright issues and on the effects of intent compromises on creative quality. Additionally, in contrast to works that directly introduced suggestions to the writers’ texts, which they wrote in the systems’ interfaces [21, 71, 86, 120, 122, 139], by providing support for writers who wrote their texts separately from our interface (in Word), our tool provided less guidance on the writing approach writers should choose, allowing greater personalization through ambiguity [39]. For instance, P8 from Section 5.1 could have planned out their entire plot by using CharacterMeet for inspiration then write without ever consulting CharacterMeet again. This greater personalization potential allows us to draw **stronger conclusions on the preferences of writers who usually wrote more linearly but adopted iterative patterns** while using our tool. Extending on participant feedback and [57]’s analysis of writing approaches, supporting greater productivity of iterative patterns, future works could delve deeper into the benefits of iterative writing processes (Section 6.2.1).

Secondly, our work is the first on a system supporting entire creative writing processes powered by GPT-4, revealing the **latest LLMs’ capabilities** on personalized idea exploration (in line with [5]’s predictions). In particular, our work establishes that **conversation with a chatbot supporting progressive manifestation of a story character can support the entire character construction process** (Section 5.2) and could allow realistic conversation with a greater variety of developed characters with advances in LLMs’ text generation capabilities. While a traditional way to categorize stories is as plot-driven or character-driven, which assumes that the character either plays a passive or an active role in the plot, some

CSI Factor	Average Factor Score (max 20)		Average Factor Count (max 5)		Average Weighted Factor Score (max 100)	
	[116]	CharacterMeet	[116]	CharacterMeet	[116]	CharacterMeet
Exploration	16.63	15.79	4.38	4.14	72.73	66.29
Expressiveness	14.63	15.93	3.50	3	51.19	48.64
Immersion	11.38	14.43	2.50	3	28.44	45.43
Enjoyment	18.13	17.50	2.13	2.07	38.52	36.93
Results worth effort	17.38	16.57	2.00	2.14	34.75	33.93
Collaboration	10.50	14.00	0.50	0.64	5.25	8.86
Overall CSI					76.96	80.02

Table 4: Results for the Creativity Support Index (CSI) questionnaire [19] for [116]’s final prototype and CharacterMeet. The factor score is the sum of scores for a pair of 10-point Likert agreement statements on user experience related to the factor. After the agreement statements, the respondent answers comparative questions assessing the relative importance of all factors by pairing them. The factor count is the number of times a factor has been chosen over another. The weighted factor score is the product of the factor score and the corresponding factor count for a respondent. The CSI score is the sum of all weighted factor scores divided by three (for a score out of 100).

scholars reject this view, suggesting a bidirectional relationship between plot and character in all stories [131]. Extending on this view and our findings, where writers use character responses under specific contexts to iteratively push forward the plot (Sections 5.1 and 5.2), such **conversations could potentially be used to support entire story writing processes**. Studies could explore generalizability with writing tasks less focused on the exposition of a specific character. Moreover, our convergence of usage/thinking patterns (to iterative) despite varying languages used for system interaction and task response writing (Section 4.1) suggests that, depending on the LLMs’ performance across languages, similar proficiency for the same language (imposed/emphasized on by prior LLM creative writing support works, e.g., [21, 71, 122, 132]) does not have to be a participant selection criterion for observing generalizable usage patterns. Specifically, our lack of mentioned language-specific inaccuracies (for English and Mandarin Chinese) among a diverse participant group builds confidence in **multicultural multilingual LLM creative writing support studies in the near future**, which could provide much needed research insights on diversity [72]. Our findings also encourage exploration of LLM performance in other languages and on cultural nuances (Section 5.4.2). Reported language inaccuracies for the Stable Diffusion generator (Section 4.1) warrant further studies, possibly on fine-tuning or on translating multilingual input to English through LLMs, for complementing LLM tools similar to ours.

Thirdly, **our work is the first that presents findings on user experience of an LLM-powered creative writing support tool that provides personalized visualization of a conversation with the writer’s own story character through both visuals and audio**, namely customizable character avatar, AI-generated context-relevant background images, and customizable character voice. Prior work on AI-powered chatbots either did not contain such visuals or audio or did not support their customization to the writer’s own character. Our findings not only reveal writers’ varying preferences for their customization (Section 5.3), providing future design insights, but also support the complementary effects of visuals and audio to thinking patterns during iterative

creative writing processes. Specifically, studies suggest that writers iteratively create and combine **mental imagery** (representations and sensory information that are voluntarily conjured without a “direct” external stimuli [102]) in new ways from memories to create their writing [25, 31, 74]. Such combination of mental representations is facilitated by visualization of ideas through CharacterMeet, with participants’ diverse visualization needs for visuals and audio aligning with the diversity in types (e.g., images and other sensory information) and degrees of vividness of mental imagery found in prior works [15, 25, 31, 36, 64, 102, 107, 121]. **This highlights the necessity for personalizable visuals and audio features for chatbot avatar creative writing support tools**, which we cover in Section 6.2.2.

6.2 Design Implications

We present design implications for leveraging chatbot avatars to support creative writing.

6.2.1 Study Design for Entire Writing Processes. Given individualized processes and variations during tool usage, studies should cover entire writing processes (or character construction processes) and compare them to writers’ usual processes for comprehensive insights. For comparability, researchers could include CSI questionnaires. More on personalizable system evaluation is discussed in Section 6.3.

6.2.2 Complementary Role to Varying Writing Processes. Chatbot avatars should support but not enforce iterative processes, possibly by allowing the writer to write in a different interface. Such interface separation could also ensure productivity despite longer generation waiting times (Section 3.4) through multitasking, as observed with participants and by prior work on creativity support [129]. For iterative creative processes, based on key principles [25, 30, 31, 57, 74, 115], similar chatbot avatars should support the **tracking and updating of ideas (both textual and non-textual) throughout entire creative processes as creators/writers continuously revisit and build on accumulated ideas**. For instance, similar systems can integrate a hide-able window separate from

the conversation (similar to *Character Description*) with fields that support the iterative updating of abstract ideas related to the notion of the character [131], the conversation context, and other customization choices (e.g., visuals and audio). As indicated by those who gave lower scores for “Exploration” (Section 5.5), partly on information tracking, future similar systems should consider the addition of a chat history organized in different ways (e.g., based on the story or the character). While a history was suggested by [116], we favored diminishing learning cost of our system that integrates novel interaction with various sources of inspiration. Similar chatbot avatars should also support **evolving exploration needs as writers progress through their writing** [30, 131]. To do so, such systems can integrate language models that do not restrict input format and content (e.g., conversation or requests for suggestions) and can generate output reflective of varied contexts and characters [5, 116] (e.g., the latest LLM). To improve impersonation quality, future systems can consider fine-tuning models with knowledge banks (Sections 3.4 and 5.4.2). For visuals and audio, future systems can allow greater freedom of exploration through text-to-3D avatar generation and neural speech synthesis [26, 60]. In addition, future systems should allow customization of art styles, the visibility of the avatar’s different parts, and the visibility of the background across stages. They should also have an option for dynamically changing backgrounds and an option to have multiple avatars conversing (Section 5.3).

6.3 Limitations

We consider study limitations. One challenge for personalizable creative writing support tool research is balancing personalization to individualized criteria with objective evaluation of outputs (e.g., task responses and system characters), which can provide stronger conclusions. After consultation with external creative writing research experts and prior work suggested a need for further research on evaluation criteria, for a first study, we decided to focus on discovering patterns that could encourage such research. To discover more objective criteria, future research could establish user profiles based on our findings (e.g., Section 5.3 for audio and visual preferences) and writing motives, which can be educational (e.g., satisfaction of a specific grading rubric), commercial (e.g., reader enjoyment [32] or similarity to a writer’s style [3, 59, 124]), and personal (more individualized). With more participants, research could also validate the number of adjustments before satisfaction as a quality measure for similar personalizable LLM chatbot features (Section 3.4). For further evaluation of similar systems’ effects on creativity and/or writing quality, readers familiar with the target author’s style could evaluate similarity between styles for task responses. Our consultation also supports external judging by writers or readers for longer studies, partly because prompt engineering includes trial and error. Additionally, a diverse larger sample size could reveal additional trends (e.g., model social biases, LLM and image generator multilingual performance, and participant AI proficiency, learning, and acceptance) and lead to stronger conclusions (e.g., through more detailed statistical analysis of CSI scores). Future studies could also give greater task choice freedom, using longer writing tasks for more comprehensive insights on characterization [130, 140], including other forms of storytelling (Section 5.4.2), or

observing participants from diverse backgrounds in the wild [111]. Future studies could include other types of character visuals (e.g., animals) in both 3D and 2D representations (with different levels of visibility), mitigating possible limiting effects on creativity of having human-like 3D character visuals only. Despite all these limitations, by revealing a convergence of thinking and usage patterns among diverse participants and the capabilities of the latest generative AI (Section 6.1), our study serves as a starting point that presents an optimistic outlook on and directions for personalizable multilingual creative writing support chatbot research that combines both textual and non-textual sources of inspiration powered by generative AI.

7 CONCLUSION

We investigated how conversation with a chatbot avatar embodying a story character could support the entire character construction process in creative writing. To do so, we examined thinking and usage patterns of 14 creative writers when writing with CharacterMeet, our prototype system that enables writers to progressively manifest a character through conversation while customizing context, avatar appearance, background image, and voice. Findings suggest that CharacterMeet facilitates entire iterative character construction processes, where participants, even those who had more linear usual approaches, alternated between writing and various use cases of character exploration through personalized visualization on CharacterMeet, with visuals and audio enhancing immersion. Our findings support research on iterative creative processes and the growing potential of personalizable LLM creativity support tools that integrate visuals and audio for culturally and linguistically diverse users. Future works could delve deeper into the benefits of AI chatbot avatars creative writing support tools that support but not enforce iterative patterns while studying entire writing processes instead of specific stages.

ACKNOWLEDGMENTS

This research was partly supported by a grant from the Guangzhou Municipal Nansha District Science and Technology Bureau under Contract No.2022ZD01.

CharacterMeet (shown in Figures 1, 4, 7a, 7e, and 7f) is powered by OpenAI GPT-4 [98] (text responses), Ready Player Me [83] (avatar visuals), and Blockade Labs’ Skybox AI [69] (background visuals). Figures 2, 7b, and 7c also show responses generated by OpenAI’s GPT-4 [98]. Figure 3 was generated using the August 2023 version of Qualtrics [105]. Figures 5 and 6 were created with draw.io [75]. We are not funded by the mentioned platforms.

We are grateful for all resources made available to us, our reviewers’ insightful feedback, and our participants’ contribution.

REFERENCES

- [1] Inworld AI. 2023. *Inworld - The most advanced Character Engine for AI NPCs*. Retrieved August 24, 2023 from <https://www.inworld.ai/>
- [2] Safinah Ali and Devi Parikh. 2021. Telling Creative Stories Using Generative Visual Aids. (2021). <https://doi.org/10.48550/arXiv.2110.14810> arXiv:arXiv:2110.14810
- [3] Vikas Ganjigunte Ashok, Song Feng, and Yejin Choi. 2013. Success with Style: Using Writing Style to Predict the Success of Novels. In *Conference on Empirical Methods in Natural Language Processing*.
- [4] Jekaterina Belakova and Wendy E. Mackay. 2021. SonAmi: A Tangible Creativity Support Tool for Productive Procrastination. In *Proceedings of the 13th Conference*

- on *Creativity and Cognition* (Virtual Event, Italy) (C&C '21). Association for Computing Machinery, New York, NY, USA, Article 7, 10 pages. <https://doi.org/10.1145/3450741.3465250>
- [5] Oloff C. Biermann, Ning F. Ma, and Dongwook Yoon. 2022. 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* (Virtual Event, Australia) (DIS '22). Association for Computing Machinery, New York, NY, USA, 1209–1227. <https://doi.org/10.1145/3532106.3533506>
 - [6] Jay David Bolter and Michael Joyce. 1987. Hypertext and creative writing. In *Proceedings of the ACM Conference on Hypertext* (Chapel Hill, North Carolina, USA) (HYPERTEXT '87). Association for Computing Machinery, New York, NY, USA, 41–50. <https://doi.org/10.1145/317426.317431>
 - [7] Ali Borji and Mehrdad Mohammadian. 2023. Battle of the Wordsmiths: Comparing ChatGPT, GPT-4, Claude, and Bard. (June 2023). <https://doi.org/10.2139/ssrn.4476855>
 - [8] Blake Brittain. 2023. *More writers sue OpenAI for copyright infringement over AI training*. Retrieved December 5, 2023 from <https://www.reuters.com/technology/more-writers-sue-openai-copyright-infringement-over-ai-training-2023-09-11/>
 - [9] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. (2023). <https://doi.org/10.48550/arXiv.2303.12712> arXiv:arXiv:2303.12712
 - [10] Vivien Burr. 2002. *The Person in Social Psychology* (1st ed.). Psychology Press, London. <https://doi.org/10.4324/9780203989173>
 - [11] Janet Burroway, Elizabeth Stuckey-French, and Ned Stuckey-French. 2019. *Writing Fiction: A Guide to Narrative Craft* (10th ed.). University of Chicago Press. <https://doi.org/10.7202/chicago/9780226616728.001.0001>
 - [12] Gabriel S. Cabrera. 2023. *BanterBot*. Retrieved August 24, 2023 from <https://pypi.org/project/BanterBot/>
 - [13] Janelynn Camingue, Elin Carstensdottir, and Edward F. Melcer. 2021. What is a Visual Novel? *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 285 (Oct. 2021), 18 pages. <https://doi.org/10.1145/3474712>
 - [14] Michael D. Carey, Shelley Davidow, and Paul Williams. 2022. Re-imagining narrative writing and assessment: a post-NAPLAN craft-based rubric for creative writing. *The Australian Journal of Language and Literacy* 45 (2022), 33–48. <https://doi.org/10.1007/s44020-022-00004-4>
 - [15] Robyn Carston. 2018. Figurative Language, Mental Imagery, and Pragmatics. *Metaphor and Symbol* 33, 3 (2018), 198–217. <https://doi.org/10.1080/10926488.2018.1481257>
 - [16] Seymour Chatman. 1972. On the Formalist-Structuralist Theory of Character. *Journal of Literary Semantics* 1, Jahresband (1972), 57–79. <https://doi.org/10.1515/jlse.1972.1.1.57>
 - [17] Seymour Chatman. 1980. *Story and Discourse: Narrative Structure in Fiction and Film*. Cornell University Press, Ithaca, NY. <https://doi.org/10.1515/9781501741616>
 - [18] Myra Cheng, Esin Durmus, and Dan Jurafsky. 2023. Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models. (2023). <https://doi.org/10.48550/arXiv.2305.18189> arXiv:2305.18189
 - [19] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (June 2014), 25 pages. <https://doi.org/10.1145/2617588>
 - [20] Bhavya Chopra, Khushali Verma, Sonali Singhal, and Utsav Singla. 2021. Reality Tales: Facilitating User-Character Interaction with Immersive Storytelling (CHI EA '21). Association for Computing Machinery, New York, NY, USA, Article 489, 7 pages. <https://doi.org/10.1145/3411763.3451522>
 - [21] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (, New Orleans, LA, USA), (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 209, 19 pages. <https://doi.org/10.1145/3491102.3501819>
 - [22] Lynda Clark and Divij Sood. 2022. Working Backwards: Creating a Character Backstory Generation System Using Idealized Creative Writing Outputs: Creating a Character Backstory Generation System Using Idealized Creative Writing Outputs. In *Proceedings of the 17th International Conference on the Foundations of Digital Games* (Athens, Greece) (FDG '22). Association for Computing Machinery, New York, NY, USA, Article 29, 9 pages. <https://doi.org/10.1145/3555858.3555867>
 - [23] Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The Journal of Positive Psychology* 12, 3 (2017), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>
 - [24] Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. 2023. Choice Over Control: How Users Write with Large Language Models using Diegetic and Non-Diegetic Prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 408, 17 pages. <https://doi.org/10.1145/3544548.3580969>
 - [25] Susan Daniels-McGhee and Gary A. Davis. 1994. The Imagery-Creativity Connection. *The Journal of Creative Behavior* 28, 3 (1994), 151–176. <https://doi.org/10.1002/j.2162-6057.1994.tb01189.x>
 - [26] Min Ding. 2023. A Systematic Review on the Development of Speech Synthesis. In *2023 8th International Conference on Computer and Communication Systems (ICCCS)*. 28–33. <https://doi.org/10.1109/ICCCS57501.2023.10150729>
 - [27] Lajos Egri. 1960. *The Art Of Dramatic Writing: Its Basis in the Creative Interpretation of Human Motives* (revised ed.). Simon & Schuster.
 - [28] Syd Field. 2006. *The Screenwriter's Workbook: Exercises and Step-by-Step Instructions for Creating a Successful Screenplay, Newly Revised and Updated*. Delta.
 - [29] David Fishelov. 1990. Types of Character, Characteristics of Types. *Style* 24, 3 (1990), 422–439.
 - [30] Linda Flower and John R. Hayes. 1981. A Cognitive Process Theory of Writing. *College Composition and Communication* 32, 4 (Dec. 1981), 365–387.
 - [31] Linda Flower and John R. Hayes. 1984. Images, Plans, and Prose: The Representation of Meaning in Writing. *Written Communication* 1, 1 (1984), 120–160. <https://doi.org/10.1177/0741088384001001006>
 - [32] Sven Form, Judith Aue, and Christian Kaernbach. 2019. Judging Popular Novels as Creative Products: Which Creative Attributes Contribute to their Success? *Creativity, Theories – Research – Applications* 6, 1 (2019), 20–41. <https://doi.org/10.1515/ctra-2019-0002>
 - [33] Edward Morgan Forster. 1974. *Aspects of the novel, and related writings* (abinger ed.). Hodder Arnold.
 - [34] Lita Furby. 1978. Possession in humans: An exploratory study of its meaning and motivation. *Social Behavior and Personality: an international journal* 6, 1 (1978), 49–65. <https://doi.org/10.2224/sbp.1978.6.1.49>
 - [35] David Galef. 1993. *The Supporting Cast: A Study of Flat and Minor Characters* (1st ed.). Penn State University Press.
 - [36] Francis Galton. 1880. Statistics of Mental Imagery. *Mind* 5, 19 (July 1880), 301–318. <https://doi.org/10.1093/mind/os-V.19.301>
 - [37] Eduardo C. Garrido-Merchán, José Luis Arroyo-Barrigüete, and Roberto Gozalo-Brihuea. 2023. Simulating H.P. Lovecraft horror literature with the Chat-GPT large language model. (2023). <https://doi.org/10.48550/arXiv.2305.03429> arXiv:arXiv:2305.03429
 - [38] James Garvey. 1978. Characterization in narrative. *Poetics* 7, 1 (1978), 63–78. [https://doi.org/10.1016/0304-422X\(78\)90005-0](https://doi.org/10.1016/0304-422X(78)90005-0)
 - [39] William W. Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) (CHI '03). Association for Computing Machinery, New York, NY, USA, 233–240. <https://doi.org/10.1145/642611.642653>
 - [40] Katy Ilonka Gero, Tao Long, and Lydia B. Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 245, 15 pages. <https://doi.org/10.1145/3544548.3580782>
 - [41] Carlos Gómez-Rodríguez and Paul Williams. 2023. A Confederacy of Models: a Comprehensive Evaluation of LLMs on Creative Writing. (2023). <https://doi.org/10.48550/arXiv.2310.08433> arXiv:arXiv:2310.08433
 - [42] Frederica Gonçalves, Diogo Cabral, Pedro Campos, and Johannes Schöning. 2017. I Smell Creativity: Exploring the Effects of Olfactory and Auditory Cues to Support Creative Writing Tasks. In *Human-Computer Interaction - INTERACT 2017*. Springer International Publishing, Cham, 165–183. https://doi.org/10.1007/978-3-319-67684-5_11
 - [43] Frederica Gonçalves, Diogo Cabral, and Pedro Campos. 2018. CreaSenses: Fostering Creativity Through Olfactory Cues. In *Proceedings of the 36th European Conference on Cognitive Ergonomics* (Utrecht, Netherlands) (ECCE '18). Association for Computing Machinery, New York, NY, USA, Article 12, 4 pages. <https://doi.org/10.1145/3232078.3232090>
 - [44] Frederica Gonçalves and Pedro Campos. 2018. Mild Place Illusion: A Virtual Reality Factor to Spark Creativity in Writing. In *Proceedings of the 36th European Conference on Cognitive Ergonomics* (Utrecht, Netherlands) (ECCE '18). Association for Computing Machinery, New York, NY, USA, Article 7, 8 pages. <https://doi.org/10.1145/3232078.3232085>
 - [45] Frederica Gonçalves, Ana Caraban, Evangelos Karapanos, and Pedro Campos. 2017. What Shall I Write Next? Subliminal and Supraliminal Priming as Triggers for Creative Writing. In *Proceedings of the European Conference on Cognitive Ergonomics* (Umeå, Sweden) (ECCE '17). Association for Computing Machinery, New York, NY, USA, 77–84. <https://doi.org/10.1145/3121283.3121294>
 - [46] Erik E. Guzik, Christian Byrge, and Christian Gilde. 2023. The originality of machines: AI takes the Torrance Test. *Journal of Creativity* 33, 3 (2023), 100065. <https://doi.org/10.1016/j.jjoc.2023.100065>
 - [47] Jason Hamilton. 2022. *17 Character Development Exercises for Writers*. Retrieved August 24, 2023 from <https://kindpreneur.com/character-development-exercises/>
 - [48] William John Harvey. 1965. *Character and the Novel*. Chatto & Windus, London.
 - [49] Steven J. Heine and Matthew B. Ruby. 2010. Cultural psychology. *Wiley Interdisciplinary Reviews: Cognitive Science* 1, 2 (March 2010), 254–266.

- [50] Baruch Hochman. 1985. *Character in Literature*. Cornell Univ Pr.
- [51] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2023. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. (2023). <https://doi.org/10.48550/arXiv.2311.05232> arXiv:arXiv:2311.05232
- [52] Shinya Iizuka, Shota Mochizuki, Atsumoto Ohashi, Sanae Yamashita, Ao Guo, and Ryuichiro Higashinaka. 2023. Clarifying the Dialogue-Level Performance of GPT-3.5 and GPT-4 in Task-Oriented and Non-Task-Oriented Dialogue Systems. In *Proceedings of the 2023 AAAI Fall Symposia*, Vol. 2. <https://doi.org/10.1609/aaais.v21i1.27668>
- [53] Character Technologies Inc. 2023. *character.ai*. Retrieved August 24, 2023 from <https://beta.character.ai/>
- [54] Convai Technologies Inc. 2023. *Convai - Conversational AI for Virtual Worlds*. Retrieved August 24, 2023 from <https://convai.com/>
- [55] Steve Ince. 2009. *Writing for Video Games* (1st ed.). Bloomsbury UK.
- [56] Tal Ivancovsky, Simone Shamy-Toory, Joo Lee, Hiroaki Morio, and Jenny Kurman. 2019. A dual process model of generation and evaluation: A theoretical framework to examine cross-cultural differences in the creative process. *Personality and Individual Differences* 139 (2019), 60–68. <https://doi.org/10.1016/j.paid.2018.11.012>
- [57] Annie Jackson. 2022. The Advantage of an Iterative Writing Process for Novels and Short Stories. *Masters Theses* (2022).
- [58] Hang Jiang, Xijie Zhang, Xubo Cao, and Jad Kabbara. 2023. PersonalLLM: Investigating the Ability of Large Language Models to Express Big Five Personality Traits. (2023). <https://doi.org/10.48550/arXiv.2305.02547> arXiv:arXiv:2305.02547
- [59] Jian Jin and Siyun Liu. 2022. An Analysis of the Linguistic Features of Popular Chinese Online Fantasy Novels. *Discourse Processes* 59, 4 (2022), 326–344. <https://doi.org/10.1080/0163853X.2022.2028432>
- [60] Navdeep Kaur and Parminder Singh. 2023. Conventional and contemporary approaches used in text to speech synthesis: a review. *Artificial Intelligence Review* 56 (2023), 5837–5880. <https://doi.org/10.1007/s10462-022-10315-0>
- [61] Lauren Kessler. 2023. *Character: The beating heart of narrative*. Retrieved August 24, 2023 from <https://niemanstoryboard.org/stories/profiles-character-development-reporting-sources/>
- [62] Kathrin Knutzen, Florian Weidner, and Wolfgang Broll. 2019. Talk to me! exploring stereoscopic 3D anthropomorphic virtual assistants in automated vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (Utrecht, Netherlands) (*AutomotiveUI '19*). Association for Computing Machinery, New York, NY, USA, 363–368. <https://doi.org/10.1145/3349263.3351503>
- [63] David Koppelman and Scott Imig. 1995. The Effect of Music on Children's Writing Content. (1995).
- [64] Stephen M. Kosslyn and Pierre Jolicoeur. 1980. A Theory-Based Approach to the Study of Individual Differences in Mental Imagery. In *Aptitude, Learning, and Instruction: Volume 2* (1st ed.). Taylor & Francis Group. <https://doi.org/10.4324/9781003163145-19>
- [65] Milan Kundera. 1988. *The Art of the Novel*. Harper & Row.
- [66] Joel J. Kupperman. 1991. *Character*. Oxford University Press.
- [67] Joung Huem Kwon, John Powell, and Alan Chalmers. 2013. How level of realism influences anxiety in virtual reality environments for a job interview. *International Journal of Human-Computer Studies* 71, 10 (2013), 978–987. <https://doi.org/10.1016/j.ijhcs.2013.07.003>
- [68] Purdue Online Writing Lab. 2024. *Writing Compelling Characters*. Retrieved February 1, 2024 from https://owl.purdue.edu/owl/subject_specific_writing/creative_writing/characters_and_fiction_writing/writing_compelling_characters.html
- [69] Blockade Labs. 2023. *Blockade Labs Skybox - AI-Generated 3D Worlds*. Retrieved August 24, 2023 from <https://skybox.blockadelabs.com/>
- [70] Latitude. 2023. *AI Dungeon*. Retrieved August 24, 2023 from <https://play.aidungeon.io/>
- [71] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 388, 19 pages. <https://doi.org/10.1145/3491102.3502030>
- [72] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 143, 14 pages. <https://doi.org/10.1145/3411764.3445488>
- [73] Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. 2023. G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment. (2023). <https://doi.org/10.48550/arXiv.2303.16634> arXiv:arXiv:2303.16634
- [74] Shirley Long and Elfrieda H. Hiebert. 1985. Effects of awareness and practice in mental imagery on creative writing of gifted children. In *Issues in Literacy: A research perspective (34th Yearbook of the National Reading Conference)*, J. Niles and R. Lalik (Eds.), 381–385.
- [75] JGraph Ltd. 2023. *draw.io*. Retrieved September 1, 2023 from <https://www.drawio.com/>
- [76] Celina L. Ma, Haohong Wang, and Mea Wang. 2023. Guiding Interactive Film With Emotion-Profiling Chatbots. In *2023 International Conference on Computing, Networking and Communications (ICNC)*, 478–483. <https://doi.org/10.1109/ICNC57223.2023.10074513>
- [77] Karl F. MacDorman, Robert D. Green, Chin-Chang Ho, and Clinton T. Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior* 25, 3 (2009), 695–710. <https://doi.org/10.1016/j.chb.2008.12.026> Including the Special Issue: Enabling elderly users to create and share self authored multimedia content.
- [78] David Mal, Erik Wolf, Nina Döllinger, Mario Botsch, Carolin Wienrich, and Marc Erich Latoschik. 2022. Virtual Human Coherence and Plausibility - Towards a Validated Scale. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 788–789. <https://doi.org/10.1109/VRW55335.2022.00245>
- [79] Mary Lee Marksberry. 1963. *Foundation of Creativity*. Harper & Row, New York.
- [80] Sybil Marshall. 1974. *Creative Writing*. Macmillan.
- [81] Marcel Marti, Jodok Vieli, Wojciech Witoń, Rushit Sanghrajka, Daniel Inversini, Diana Wotruba, Isabel Simo, Sasha Schriber, Mubbasir Kapadia, and Markus Gross. 2018. CARDINAL: Computer Assisted Authoring of Movie Scripts. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) (*IUI '18*). Association for Computing Machinery, New York, NY, USA, 509–519. <https://doi.org/10.1145/3172944.3172972>
- [82] Ready Player Me. 2024. *Build your own Avatar Creator*. Retrieved February 1, 2024 from <https://docs.readyplayer.me/ready-player-me/integration-guides/unity/avatar-creator/build-your-own-avatar-creator>
- [83] Ready Player Me. 2024. *Licensing & Privacy*. Retrieved February 1, 2024 from <https://docs.readyplayer.me/ready-player-me/support/terms-of-use>
- [84] Rick Meyer. 2005. *14 Tips for Building Character*. Retrieved August 24, 2023 from <https://niemanstoryboard.org/stories/14-tips-for-building-character/>
- [85] Paul Mills. 2005. *The Routledge Creative Writing Coursebook* (1st ed.). Routledge.
- [86] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 355, 34 pages. <https://doi.org/10.1145/3544548.3581225>
- [87] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- [88] David Morley. 2007. *The Cambridge Introduction to Creative Writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511803024>
- [89] Michael W. Morris and Kwok Leung. 2010. Creativity East and West: Perspectives and Parallels. *Management and Organization Review* 6, 3 (2010), 313–327. <https://doi.org/10.1111/j.1740-8784.2010.00193.x>
- [90] Hamideh Mozaffari. 2013. An Analytical Rubric for Assessing Creativity in Creative Writing. *Theory & Practice in Language Studies* 3, 12 (Dec. 2013), 2214–2219. <https://doi.org/10.4304/tpls.3.12.2214-2219>
- [91] Donald M. Murray. 1991. All Writing Is Autobiography. *College Composition and Communication* 42, 1 (Feb. 1991), 66–74. <https://doi.org/10.2307/357540>
- [92] Scott Myers. 2022. *The Protagonist's Journey: An Introduction to Character-driven Screenwriting and Storytelling* (1st ed.). Palgrave Macmillan Cham. <https://doi.org/10.1007/978-3-030-79682-2>
- [93] Roberto Navigli, Simone Conia, and Björn Ross. 2023. Biases in Large Language Models: Origins, Inventory, and Discussion. *J. Data and Information Quality* 15, 2, Article 10 (June 2023), 21 pages. <https://doi.org/10.1145/3597307>
- [94] Justin Nicholes. 2017. Measuring ownership of creative versus academic writing: Implications for interdisciplinary praxis. *Writing in Practice* 3 (2017).
- [95] OpenAI. 2023. *GPT-4*. Retrieved August 24, 2023 from <https://openai.com/research/gpt-4>
- [96] OpenAI. 2023. *GPT-4 Technical Report*. arXiv:2303.08774 [cs.CL]
- [97] OpenAI. 2023. *OpenAI Platform*. Retrieved August 24, 2023 from <https://platform.openai.com/docs/api-reference>
- [98] OpenAI. 2023. *Usage policies*. Retrieved August 24, 2023 from <https://openai.com/policies/usage-policies>
- [99] Hiroyuki Osone, Jun-Li Lu, and Yoichi Ochiai. 2021. BunCho: AI Supported Story Co-Creation via Unsupervised Multitask Learning to Increase Writers' Creativity in Japanese. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI EA '21*). Association for Computing Machinery, New York, NY, USA, Article 19, 10 pages. <https://doi.org/10.1145/3411763.3450391>
- [100] Pat Pataranutaporn, Valdemar Danry, Lancelot Blanchard, Lavanay Thakral, Naoki Ohsugi, Pattie Maes, and Misha Sra. 2023. Living Memories: AI-Generated Characters as Digital Mementos. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) (*IUI '23*). Association for Computing Machinery, New York, NY, USA, 889–901. <https://doi.org/10.1145/3581641.3584065>

- [101] Joanna Pawliczak. 2015. Creative Writing as a Best Way to Improve Writing Skills of Students. *Sino-US English Teaching* 12, 5 (May 2015), 347–352. <https://doi.org/10.17265/1539-8072/2015.05.004>
- [102] Joel Pearson. 2019. The human imagination: the cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience* 20, 10 (2019), 624–634. <https://doi.org/10.1038/s41583-019-0202-9>
- [103] James Phelan. 1989. *Reading People, Reading Plots: Character, Progression, and the Interpretation of Narrative*. University of Chicago Press.
- [104] Eva Pietroni, Massimiliano Forlani, and Claudio Rufa. 2015. Livia's Villa Reloaded: An example of re-use and update of a pre-existing Virtual Museum, following a novel approach in storytelling inside virtual reality environments. In *2015 Digital Heritage*, Vol. 2. 511–518. <https://doi.org/10.1109/DigitalHeritage.2015.7419567>
- [105] Qualtrics. 2023. *Qualtrics XM: The Leading Experience Management Software*. Retrieved September 1, 2023 from <https://www.qualtrics.com>
- [106] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. (2019).
- [107] Alan Richardson. 1969. *Mental Imagery*. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-37817-5>
- [108] Shlomith Rimmon-Kenan. 2002. *Narrative Fiction: Contemporary Poetics* (2nd ed.). Routledge, London. <https://doi.org/10.4324/9780203426111>
- [109] Edgar V. Roberts and Henry E. Jacobs. 1998. *Literature: An Introduction to Reading and Writing* (5th ed.). Prentice Hall.
- [110] Alicia Rodriguez. 2008. The 'Problem' of Creative Writing: Using Grading Rubrics Based On Narrative Theory as Solution. *New Writing* 5, 3 (2008), 167–177. <https://doi.org/10.1080/14790720802209963>
- [111] Yvonne Rogers and Paul Marshall. 2017. *Research in the Wild* (1st ed.). Springer, Cham. <https://doi.org/10.1007/978-3-031-02220-3>
- [112] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. (2022). <https://doi.org/10.48550/arXiv.2112.10752> arXiv:arXiv:2112.10752
- [113] Pegah Salehi, Syed Zohaib Hassan, Saeed Shafiee Sabet, Gunn Astrid Baugerud, Miriam Sinkerd Johnson, Pål Halvorsen, and Michael A. Riegler. 2022. Is More Realistic Better? A Comparison of Game Engine and GAN-Based Avatars for Investigative Interviews of Children. In *Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval* (Newark, NJ, USA) (ICDAR '22). Association for Computing Machinery, New York, NY, USA, 41–49. <https://doi.org/10.1145/3512731.3534209>
- [114] Angela Salmon. 2010. Using music to promote children's thinking and enhance their literacy development. *Early Child Development and Care* 180, 7 (2010), 937–945. <https://doi.org/10.1080/03004430802550755>
- [115] R. Keith Sawyer. 2021. The iterative and improvisational nature of the creative process. *Journal of Creativity* 31 (2021), 100002. <https://doi.org/10.1016/j.joc.2021.100002>
- [116] Oliver Schmitt and Daniel Buschek. 2021. 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* (Virtual Event, Italy) (C&C '21). Association for Computing Machinery, New York, NY, USA, Article 10, 10 pages. <https://doi.org/10.1145/3450741.3465253>
- [117] Guy Schofield, Gareth Beale, Nicole Beale, Martin Fell, Dawn Hadley, Jonathan Hook, Damian Murphy, Julian Richards, and Lewis Thresh. 2018. Viking VR: Designing a Virtual Reality Experience for a Museum. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) (DIS '18). Association for Computing Machinery, New York, NY, USA, 805–815. <https://doi.org/10.1145/3196709.3196714>
- [118] Daniel R. Schwarz. 1990. Character and Characterization: An Inquiry. In *The Case For a Humanistic Poetics*. Palgrave Macmillan UK, London, 65–104. https://doi.org/10.1007/978-1-349-11070-4_3
- [119] Linda Seger. 1990. *Creating Unforgettable Characters: A Practical Guide to Character Development in Films, TV Series, Advertisements, Novels & Short Stories* (1st ed.). Holt Paperbacks.
- [120] Hanieh Shakeri, Carman Neustaedt, and Steve DiPaola. 2021. SAGA: Collaborative Storytelling with GPT-3. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) (CSCW '21). Association for Computing Machinery, New York, NY, USA, 163–166. <https://doi.org/10.1145/3462204.3481771>
- [121] Gilbert Simondon. 2023. *Imagination and Invention*. Univocal Publishing.
- [122] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2023. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Transactions on Computer-Human Interaction* 30, 5, Article 68 (Sept. 2023), 57 pages. <https://doi.org/10.1145/3511599>
- [123] Yun-Zhu Song, Zhi Rui Tam, Hung-Jen Chen, Huiao-Han Lu, and Hong-Han Shuai. 2020. Character-Preserving Coherent Story Visualization. In *Computer Vision – ECCV 2020*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 18–33. https://doi.org/10.1007/978-3-030-58520-4_2
- [124] Zhivar Sourati Hassan Zadeh, Nazanin Sabri, Houmaan Chamani, and Behnam Bahrak. 2022. Quantitative analysis of fanfictions' popularity. *Social Network Analysis and Mining* 12, Article 42 (March 2022). <https://doi.org/10.1007/s13278-021-00854-9>
- [125] Sol Stein. 1995. *Stein on Writing: A Master Editor of Some of the Most Successful Writers of Our Century Shares His Craft Techniques and Strategies* (1st ed.). St. Martin's Press.
- [126] Sudowrite. 2023. *Sudowrite*. Retrieved August 24, 2023 from <https://www.sudowrite.com/>
- [127] Yuqian Sun, Xingyu Li, Jun Peng, and Ze Gao. 2023. Inspire creativity with ORIBA: Transform Artists' Original Characters into Chatbots through Large Language Model. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (Cancun, Quintana Roo, Mexico,) (UbiComp/ISWC '23 Adjunct). Association for Computing Machinery, New York, NY, USA, 78–82. <https://doi.org/10.1145/3594739.3610695>
- [128] Toolsaday. 2023. *AI Character Generator*. Retrieved August 24, 2023 from <https://toolsaday.com/writing/character-generator>
- [129] Josh Urban Davis, Fraser Anderson, Merten Stroetzl, Toví Grossman, and George Fitzmaurice. 2021. Designing Co-Creative AI for Virtual Environments. In *Proceedings of the 13th Conference on Creativity and Cognition* (Virtual Event, Italy) (C&C '21). Association for Computing Machinery, New York, NY, USA, Article 26, 11 pages. <https://doi.org/10.1145/3450741.3465260>
- [130] Maryam Vaezi and Saeed Rezaei. 2019. Development of a rubric for evaluating creative writing: a multi-phase research. *New Writing* 16, 3 (2019), 303–317. <https://doi.org/10.1080/14790726.2018.1520894>
- [131] Lina Varotsi. 2019. *Conceptualisation and Exposition: A Theory of Character Construction* (1st ed.). Routledge, New York. <https://doi.org/10.4324/9780429060762>
- [132] Qian Wan, Siying Hu, Yu Zhang, Piao Hong Wang, Bo Wen, and Zhicong Lu. 2023. "It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. (2023). <https://doi.org/10.48550/arXiv.2307.10811> arXiv:arXiv:2307.10811
- [133] Qian Janice Wang, Francisco Barbosa Escobar, Signe Lund Mathiesen, and Patricia Alves Da Mota. 2021. Can Eating Make Us More Creative? A Multisensory Perspective. *Foods* 10, 2 (Feb. 2021), 469. <https://doi.org/10.3390/foods10020469>
- [134] Xintao Wang, Quan Tu, Yaying Fei, Ziang Leng, and Cheng Li. 2023. Does Role-Playing Chatbots Capture the Character Personalities? Assessing Personality Traits for Role-Playing Chatbots. (2023). <https://doi.org/10.48550/arXiv.2310.17976> arXiv:arXiv:2310.17976
- [135] John A Waterworth. 1997. Creativity and Sensation: The Case for Synaesthetic Media. *Leonardo* 30, 4 (1997), 327–330. <https://doi.org/10.2307/1576481>
- [136] Katie Watson, Samuel S. Sohn, Sasha Schriber, Markus Gross, Carlos Manuel Muniz, and Mubbasir Kapadia. 2019. StoryPrint: An Interactive Visualization of Stories. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) (IUI '19). Association for Computing Machinery, New York, NY, USA, 303–311. <https://doi.org/10.1145/3301275.3302302>
- [137] Dominik Welke, Isaac Purton, and Edward A. Vessel. 2023. Inspired by art: Higher aesthetic appeal elicits increased felt inspiration in a creative writing task. *Psychology of Aesthetics, Creativity, and the Arts* 17, 3 (2023), 261–277. <https://doi.org/10.1037/aca0000393>
- [138] Bernard Williams. 1995. *Making Sense of Humanity: And Other Philosophical Papers 1982-1993*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511621246>
- [139] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) (IUI '22). Association for Computing Machinery, New York, NY, USA, 841–852. <https://doi.org/10.1145/3490099.3511105>
- [140] Claire M. Zedelius, Caitlin Mills, and Jonathan W. Schooler. 2019. Beyond subjective judgments: Predicting evaluations of creative writing from computational linguistic features. *Behavior Research Methods* 51, 2 (April 2019), 879–894. <https://doi.org/10.3758/s13428-018-1137-1>
- [141] Chao Zhang, Cheng Yao, Jiayi Wu, Weijia Lin, Lijuan Liu, Ge Yan, and Fangtian Ying. 2022. StoryDrawer: A Child-AI Collaborative Drawing System to Support Children's Creative Visual Storytelling. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 311, 15 pages. <https://doi.org/10.1145/3491102.3501914>