Comparative Analysis of Differentiated Approaches to Utilizing AI for Subverting Stereotypes

Xiaohan Feng [1,*] and Makoto Murakami [2]

[1] Graduate School of Information Sciences and Arts, Toyo University, Kawagoe, Saitama, Japan
[2] Dept. of Information Sciences and Arts, Toyo University, Kawagoe, Saitama, Japan
Email: nooothing2@gmail.com (X.F.); murakami_m@toyo.jp (M.M.)
*Corresponding author

***Abstract*—Limited or superficial knowledge about others can foster stereotypes and prejudice. Consequently, this study explores specific methods to counteract these stereotypes. We posit that the key to challenging stereotypes lies in acquiring relevant knowledge about the subject. Given that gathering such information is often time-consuming, this study introduces an Artificial Intelligence (AI) -assisted, time-efficient approach. Additionally, we outline the advantages and disadvantages of these methods and examine their suitability across different scenarios. This study has three primary objectives. First, we propose targeted methods to reduce stereotypical visual elements in current designs. These methods aim not only to create novel visual representations of characters but also to ensure they remain recognizable to the audience. Second, we seek to mitigate the negative effects associated with stereotypes. We will experimentally test the effectiveness of this method, with success indicated by the character's representation surpassing existing stereotypes. Finally, we analyze and compare experimental data to evaluate the strengths and weaknesses of the proposed methods, as well as their appropriate applications. In general, this study examines three methods of deconstructing visual stereotypes across three phases, with the data compared and summarized. The methods include: 1) analyzing textual media to deconstruct stereotypes of characters whose origins conflict with current stereotypes; 2) examining multiple media sources and historical records to address stereotypes of characters that have undergone multiple transformations over time; and 3) using ChatGPT-4 to summarize archetypal character stereotypes, then deconstructing these stereotypes to assess whether they are exchanged or retained. This study achieved three primary objectives. First, it proposed specific methods to reduce standardized visual elements, or stereotypes, in current designs. These methods generated new visual representations of the character while ensuring that the character remained recognizable to the audience. Second, the study aimed to mitigate the negative effects associated with stereotypes. Experimental data indicated that the characters used in the experiment broke from existing stereotypes. Given that it takes time for elements of current stereotypes to shift and become widely accepted as new norms, it is theoretically possible to lessen the negative impacts of existing stereotypes by incorporating the character concept into work over time, using the methodology developed in this study. The third objective was to analyze and compare the experimental data to summarize the strengths and weaknesses of the methodologies, as well as the contexts in which they are most appropriately applied. Unlike many studies, this one refrains from directly undermining prejudicial thinking; instead, it proposes multiple methodological approaches. These methodologies are intended as tools for transforming the visual stereotypes of characters. The study employs AI to summarize stereotypes, significantly reducing time demands. It also assesses the success of breaking stereotypes based on the originality and communicative clarity of the work. Overall, this study addresses the limitations and inertia of design thinking, promoting a character-based approach that encourages critical and dialectical thinking among creators. It also provides valuable reference points and inspiration for future research on stereotypes.**

***Keywords*—artificial intelligence, stereotypes, character design, image generation**

I. INTRODUCTION

The rapid development of Artificial Intelligence (AI) in recent years has also impacted stereotypes. Numerous studies on AI and prejudice demonstrate that the patterns, predictions, and behaviors suggested by AI systems are influenced by the data sets used for training. When systems are trained on real-world data, they can produce biased outcomes that discriminate against underrepresented groups. Moreover, AI often mirrors the assumptions and biases of its developers, highlighting a growing realization among researchers and developers that stereotypes and biases are embedded in some algorithms. Consequently, these algorithms are likely to propagate and reinforce existing stereotypes [1]. One study found that algorithms trained on biased data led to algorithmic discrimination [2]. An audit of performance differences in commercial face analysis models revealed that male faces were less likely to be misrecognized than female faces. Male faces with lighter skin tones had the fewest errors, while female faces with darker skin tones were misrecognized most frequently. These significant

Manuscript received on September 26, 2024, revised November 5, 2024; accepted December 9, 2024; published March 14, 2025.

classification differences, based on skin color and gender, prompted researchers to call for more equitable and transparent face analysis algorithms from commercial companies [3].

In addition, stereotypes are often visually evident in image generation results. For example, the *Washington Post* observed that the image generation tool Stable Diffusion XL produced stereotypical outcomes. When "football" was entered, the generated image typically showed a dark-skinned male athlete, while "cleaning" led to images of smiling women performing household chores. This highlighted not only gender but also racial stereotypes. Similarly, entering "Social Security beneficiaries" produced predominantly dark-skinned individuals, while "productive people" displayed mostly white males. According to the latest Census Bureau data, 63% of food stamp recipients were white and 27% were black [4], demonstrating that these image generation results do not accurately reflect reality and likely indicate biases in the data sets.

Deep learning-based computer vision is a significant AI domain; however, large datasets used in computer vision introduce challenges, such as stereotyping. The LAION-400M dataset, created from the Common-Crawl dataset and CLIP-filtered image-text pairs, was found to contain problematic content, including pornography, stereotypes, and racist material [5]. OpenAI's DALL-E 3 image generator has also been reported to exhibit a "Western bias," favoring Western perspectives with a disproportionate focus on Caucasian representations. LAION, the non-profit organization behind Stable Diffusion, acknowledged that its image generator tends to reflect a white-centric worldview. While some tech companies conceal dataset contents due to copyright or NSFW concerns, open-source projects like Stable Diffusion and LAION allow external model inspections. Stability AI, which supports these models, emphasizes model transparency as critical to eliminating stereotyping and bias. Despite these efforts, many stereotypes persist in these models [6].

AI is now widely employed to support creative work, and some educators encourage students to use image generation tools for inspiration. However, as generated images often embody stereotypes, failure to engage in critical and creative thinking during the creative process can lead to creations that simply perpetuate these stereotypes. Given this, media creators should reevaluate the role of stereotypes in the creative process.

This study acknowledges the difficulty of altering biased attitudes. Even when stereotypes are identified and countered, authoritarianism and scapegoating may lead individuals to maintain prejudices, reinforcing stereotypes. One approach to addressing stereotypes at their root is through media literacy education. Although many studies affirm this approach's effectiveness, some findings indicate that stereotypes still persist in students' creations. Moreover, the resources required for comprehensive media literacy education are significant. While this form of education has lasting value, faster methods are needed. Consequently, this study seeks not a radical change in prejudicial thinking but rather the development of practical methodologies. By using these methodologies, visual stereotypes in character representation can be altered.

Stereotypes are social cognitive phenomena that are typically stable and persistent. They can hinder human understanding, impede the learning of new concepts, and even stifle innovation. While stereotypes may be neutral, negative, or positive, numerous studies indicate that they generally have adverse effects. To foster innovation and mitigate these negative impacts, this study proposes specific methods to subvert stereotypes. One key approach is conducting foundational research on the creative subject matter, as superficial understanding or lack of knowledge about other groups often results in stereotypes and prejudices. However, such investigations typically require a significant time investment. To address this, the study introduces a time-efficient method using AI in the second and third stages of the research process.

Stereotypes stem from natural cognitive processes but are often inaccurate for two reasons. First, stereotypes do not apply universally to all individuals within a group. Second, some stereotypes are based on prejudice and are fundamentally incorrect. Thus, stereotypes lack accuracy. This study was inspired by the notion that to dismantle stereotypes, one must first address their inaccuracies. The key to rectifying inaccuracy lies in understanding the truth and acquiring more comprehensive information.

Recognizing the time-intensive nature of these investigations, this study proposes AI-based, time-saving methods. Furthermore, we outline the strengths and limitations of these methods and discuss their applicability in various contexts.

This study has three objectives. First, we propose methods for reducing standardized visual elements (stereotypes) in contemporary designs. These methods should not only create new visual representations of characters but also ensure the characters remain recognizable to the audience. The second objective is to mitigate stereotypes' negative effects. This method will undergo experimental testing, and if effective, the results will indicate a successful break from existing stereotypes. The third objective is to analyze and compare the experimental data, summarizing each method's strengths and limitations and identifying the contexts in which each is most effective.

In the first phase, we developed a method to analyze and reconstruct characters in written media, aligning them closely with their origins to challenge visual stereotypes and evaluate this approach's feasibility. The second phase broadened information sources, integrating textual, graphic, and historical data to enhance accuracy and depth. Research shows that incorporating multiple perspectives reduces stereotypes, and this phase further tested the feasibility of a diversity-focused framework. Both phases faced a common limitation: the time-intensive nature of information gathering. To address this, the third phase introduced a dual-role approach, simultaneously dismantling stereotypes of two characters.

This method proved more efficient and focused than modifying individual characters alone. This study presents several original contributions:

1. Unlike many studies, it avoids targeting prejudicial thinking at its root, instead proposing methodological approaches to change visual stereotypes of characters.
2. This study supports the idea that "gathering more information is key to breaking stereotypes." While inspired by similar frameworks, this study emphasizes the importance of investigation over uninformed diversity.
3. AI is utilized to efficiently summarize stereotypes, saving time.
4. Unique techniques like stereotype swapping and holding are introduced to disrupt stereotypes.
5. The third-stage approach allows two prototypes to dismantle stereotypes simultaneously, promoting a directional, time-efficient reciprocal exchange.
6. Success in breaking stereotypes will be measured by the originality and communicative effectiveness of the work.

Overall, this study addresses limitations in traditional design thinking, providing a framework that encourages critical and dialectical thinking for creators. We believe it holds value as a reference and inspiration for future research on stereotypes.

II. Literature Review

In cultural media, stereotypes frequently stem from dominant perspectives on minority or marginalized groups. These perspectives often embody elements of elitism, hegemony, and patriarchy and are seldom initiated by the marginalized groups themselves, who generally have limited influence. When media is controlled by dominant groups, such as social elites, it tends to reflect these groups' cultural discourses and values [7].

Research on the prevalence of stereotypes in media and popular culture has a long history, with early studies primarily focusing on portrayals of African Americans in mainstream U.S. media [8]. As scholarly interest grew, research expanded to encompass stereotypes related to various ethnic, racial, and marginalized groups, such as Asians [9], Latinos [10], Indigenous peoples [11], and the elderly [12]. In recent years, this area of study has diversified to examine stereotypes across multiple forms of media, including television, film, video games, social media platforms, and media-sharing sites.

The emergence of media stereotypes around the 1930s can largely be attributed to the commodification of film, as capitalist imperatives drove the industrialization of film production for maximum profit. This industrialized, cyclical production involved the repetitive use of certain elements, which garnered favorable public feedback and established expectations for these recurring motifs. Such repetition effectively entrenched stereotypes, which Roland aptly described as "repetition machines" [13, 14]. Over time, these stereotypes evolved into mere symbols, serving as shorthand for character identities and events. In other words, viewers became conditioned to recognize specific schemas and associate them with particular character traits, personalities, and plot developments. This internalization of stereotypes through media consumption forms a knowledge framework often referred to as "media competence." In popular culture, these stereotypes become cultural icons, shaping the imaginary worlds of both viewers and creators, who themselves are part of this cultural environment, making it challenging for them to remain unaffected by these stereotypes.

In this context, the present study proposes several strategies for creators to deliberately deconstruct stereotypes. Breaking stereotypes, as conceptualized here, involves disrupting established media knowledge structures—namely, the fixed schemas, symbols, and standardized design practices that dominate character portrayal.

Stereotypes in the media not only hinder innovation but also contribute to broader societal detriments. Research has shown that racial minorities are significantly underrepresented across media forms, including television, film, and video games. For instance, a content analysis of television shows revealed that only one in five characters belongs to a racial minority group [15]. This underrepresentation is further substantiated by studies that highlight the absence or marginalization of minority groups across different media platforms [16].

Some scholars argue that media audiences develop perceptions of the "real world" based on the content they consume, often transferring these skewed representations to real-life situations. Stereotypes are a key element of this misinformation, distorting viewers' perceptions of social reality [17]. For example, one study found that heavy television viewers were more likely than light viewers to perceive Native Americans as less open-minded [18]. Another study revealed that heavy video gamers were less likely to hold egalitarian views about Black individuals than light gamers [19].

Beyond media stereotypes, the rapid development of Artificial Intelligence (AI) has further shaped societal biases. A growing body of research examines AI's role in reinforcing stereotypes, showing that the patterns, predictions, and behaviors generated by AI systems often mirror the biases embedded in their training data. When trained on real-world data, these systems can yield biased decisions that discriminate against underrepresented groups. Additionally, algorithm developers' assumptions and biases are often embedded within AI systems, leading to the reinforcement of stereotypes [1]. AI is increasingly used in creative work, with some educators encouraging students to draw inspiration from AI-generated images. However, the output of such systems often reflects entrenched stereotypes. Without critical and creative approaches, AI-generated works risk perpetuating these stereotypes rather than challenging or subverting them.

In light of these findings, it is essential for media creators and developers to reconsider the role of stereotypes in their creative processes and strive to mitigate the propagation of biased representations.

III. MATERIALS AND METHODS

Villains play a critical role in the plot of most narratives, often driven by malicious motives or acting in direct opposition to the protagonist. Their actions typically serve as a catalyst for plot development. A study by Ollie Johnston and Frank Thomas found that 75% of Disney film villains are portrayed as older and less attractive, with only 25% depicted as thin. Similarly, Meredith Li-Vollmer and Mark E. LaPointe observed that Disney villains frequently display “gender-crossing” traits, noting that many female Disney villains possess subtly masculine features. Their facial features, body shapes, and behaviors contrast sharply with those of the traditionally feminized characters, especially in Disney’s Princess series [20]. Thus, Disney’s representation of villains reinforces the “gender-crossing” stereotype. Li-Vollmer and LaPointe [21] further argued that gender identity in media is often reinforced through socially prescribed identifiers. When these “social rules” are violated, audiences may quickly perceive the character as deviating from societal ideals.

In recent years, Disney has experimented with new approaches to villain characterization. In the 53rd animated feature *Frozen* (also known as *Anna and the Snow Queen*), Prince Hans is portrayed as equally handsome as the protagonists, exhibiting conventional male attributes. This portrayal breaks the traditional mold for Disney villains, marking a departure from established norms. However, villain stereotypes are evolving, with increasing depictions of villains as physically attractive and even likable to audiences. In these portrayals, a villain’s beauty may symbolize power or the possession of formidable strength.

Nonetheless, many male villains continue to lack the physical appearance associated with protagonists. In a study that ranked the 100 greatest movie villains of all time, researchers compared these villains with heroes [22]. Excluding non-human characters and those with unrecognizable hairstyles, the study found a 64% rate of hair loss among villains, significantly higher than that of heroes in the same films. In the top 10 “greatest films,” the hair loss rate for villains reached 90%. This raises the question of whether a villain’s hair—or lack thereof—symbolizes their immoral disposition. Kyriakou *et al*. [23] suggest that the psychological impact of this stereotype is significant, potentially reflecting societal attitudes toward hair loss. The stereotype linking hair loss with villainy may exacerbate psychological burdens for individuals with alopecia, potentially resulting in negative social consequences such as anxiety and depression. Furthermore, the association of alopecia with “evil” may influence social body image norms. Studies indicate that alopecia is linked with greater stigma, prompting some individuals to wear wigs for increased social confidence and conformity to societal norms [24].

In addition to alopecia, age-related stereotypes about villains also have notable social effects. Miller [25] indicates that children are highly susceptible to socialization through media, such as cartoons and children’s television. Commercial children’s programming frequently relies on stereotypical characters, as time constraints often necessitate shortcuts that allow audiences to quickly grasp the narrative and character intentions. Although this approach may evoke an “emotional response,” it also results in the omission of complexities that could prompt deeper reflection [26]. Falchikov *et al*. [27] found that children, when asked to describe older people, often associated them with negative traits such as unattractiveness, ill health, loneliness, or reliance on assistive devices. These perceptions align with the common portrayals of elderly characters in children’s media. Children are particularly susceptible to such stereotypes because they are in a developmental stage characterized by imitation and mimicry. In children’s animated programming, 38% of older characters are portrayed negatively [28], and even positive older characters are often depicted with unfavorable traits. Older characters cast as villains are frequently characterized as angry or mentally unstable [29]. Additionally, children’s difficulty distinguishing between fantasy and reality may reinforce prejudices against older people [30].

Villain stereotypes often perpetuate the notion that characters who are old, unattractive, or bald are more likely to be portrayed as evil or morally ambiguous. Both children and mass media play significant roles in perpetuating these stereotypes and shaping societal attitudes. When internalized, media stereotypes can solidify into persistent biases, influencing perceptions over time. Therefore, dismantling villain stereotypes is arguably more urgent than addressing stereotypes of other character types. This study focuses on three specific villain archetypes: villains portrayed as beasts, such as dragons; villains depicted as women, such as witches; and villains characterized as men, including tyrants and mad scientists.

- Phase 1: Dragon

In authoritarian frameworks, prejudice is often rationalized. When confronted with anti-prejudice or anti-stereotype ideas, individuals with authoritarian beliefs tend to reject new perspectives, adhering to their own views and reinforcing divisions between themselves and others. This rigid mindset stems from a simplistic, personal system of thought that enables prejudiced individuals to present their views with apparent conviction [31]. Secord and Bachman, on the other hand, suggest that people experiencing frustration in their daily lives may redirect it toward a scapegoat if they cannot confront the root cause. Consequently, socially marginalized or exploited groups are more likely to experience deepened stereotypes [32]. Prejudiced thinking, therefore, proves difficult to alter. Even when individuals attempt to challenge their stereotypes, authoritarian attitudes and scapegoating can perpetuate prejudices against new groups, reinforcing stereotypes further.

Media literacy education is instrumental in reducing the influence of stereotypes. By encouraging learners to critically engage with representations of gender, race, class, and power, it provides tools to question and counter biased, discriminatory, or oppressive media messages [33]. Research shows that youth aged 15–27 who receive media literacy training are better equipped to critically evaluate political content, even when it aligns with their preexisting beliefs [34]. Studies have also highlighted media literacy's potential to mitigate the impact of stereotypes [35, 36]. While media literacy is still developing in East Asia, it is more established in Europe and the United States. However, its effectiveness in dismantling stereotypes has not been as transformative as expected. In one study, college students enrolled in a media literacy course critically examined African American stereotypes through both media consumption and production. Despite demonstrating critical understanding of stereotypes and their commercial media contexts, the students, when creating their own media, continued to reproduce stereotypes, believing that audiences would recognize these depictions as satire. This finding underscores the challenge: while media literacy fosters critical engagement, deeply ingrained stereotypes can persist in creative outputs [37].

Rather than attempting to dismantle prejudiced mindsets, this study seeks to develop methodologies that can serve as tools to modify visual stereotypes in character design. Although stereotypes originate from normal cognitive processes, they are inherently inaccurate. Perry argues that stereotypes are flawed because they ignore individual differences within a group, assigning universal characteristics based solely on group membership. This generalization inevitably leads to inaccuracies, as not all individuals conform to the stereotype [38].

This study is guided by the idea that challenging a stereotype requires first addressing its inaccuracies. The key to achieving this lies in uncovering factual information and gathering more nuanced details. To begin dismantling the dragon stereotype, the study developed a method to examine the character's portrayal in early written texts. By reconstructing the character based on textual descriptions, we aimed to provide a more accurate representation that challenges contemporary visual stereotypes.

There is a notable disparity between the modern visual representation of dragons and their descriptions in early texts. Even early visual depictions do not consistently reflect textual accounts. To minimize the influence of stereotypical imagery, the study's first phase avoided historical illustrations and visual artworks, instead prioritizing textual analysis. The primary sources were the *Book of Revelation* and English translations of the Bible. Segment-based searches were conducted to examine the dragon's head, body, and other attributes as described in the *Book of Revelation*, reorganizing these elements to construct a new graphic design. This design provided the foundation for 3D modeling, and a subsequent survey evaluated the effectiveness of this approach.

In Phase 1, we identified which stereotypes to retain or modify, drawing primarily on textual information from the *Book of Revelation* and related biblical sources. A more specific documentation of the methodology of this phase can be found in our previous paper [39]. The criteria for preserving or altering stereotypes in the first phase were based on the content of the written media records. For instance, according to the Book of Revelation, the dragon is described as having "seven heads," "ten horns," a "diadem," and "a huge tail." These elements were faithfully retained in the design of the first phase.

- Phase 2: Witch

The second phase of this study highlights the importance of diversifying sources alongside the pursuit of truth and enriched information. Research in various fields has shown that diversity is a critical factor in reducing stereotypes. For instance, the transitional learning approach suggests that transferring knowledge from one domain to another strengthens the learning and application of diversity principles [40].

Zowghi *et al*. [41] discuss diversity and inclusion in AI from a sociotechnical perspective, addressing potential biases. They argue for a human-centered approach within the AI ecosystem and propose a guideline known as DI-AI, which recommends that AI developers consider a wide range of experiences, knowledge, and backgrounds. Such an approach is essential for mitigating risks and broadening perspectives. These guidelines emphasize that prioritizing diversity within data science teams is vital for achieving more balanced and inclusive outcomes.

In this phase, information is drawn from textual media, graphic media, and historical stereotypical images, adhering to the principle of factual diversity. This information will serve as both evidence and inspiration to promote diversity in design. Specifically, AI-based image generation will be used to compile contemporary stereotypical images of witches. Following this, stereotypical images of witch accessories and appearances will be analyzed from historical and social perspectives to understand the factors shaping these representations over time. This analysis will inform a design methodology that extracts, preserves, and transforms historical stereotypes, ultimately leading to the development of a new, reimagined design.

At this stage, AI technology is introduced to facilitate a deeper understanding of current stereotypical elements within the character. AI also offers structured guidelines for examining each element in detail. The aim is to explore diverse sources and link past stereotypes with the present in order to dismantle existing biases. In Phase 2, we focus on identifying which stereotypes to retain or modify, primarily by examining historical information about witches to highlight the differences between past and present stereotypes. A more specific documentation of the methodology of this phase can be found in our previous paper [42]. The criteria for preserving or modifying stereotypes in the second phase were informed by diverse historical accounts of witches, with particular emphasis on how their image has evolved over time. Adjustments were made to historical stereotypes that deviate from contemporary perceptions. For instance, while the predominant stereotype of witches in popular culture is female, historical records reveal a significant proportion of male witches. To reflect this, the male witch element was incorporated into the design. Conversely, the broomstick element was retained, as its association with witches has remained largely unchanged since the 15th century.

- Phase 3: Tyrant and Mad Scientist, Explorer and AI Assistant

In the study aimed at mitigating stereotypes, researchers initially used post-processing techniques to adjust portraits and enhance diversity, which inspired the third phase of the research. This phase focused on modifying existing designs to subvert stereotypes, emphasizing informed investigation over mere diversification. A common challenge in the first and second phases was the extensive time required for thorough information gathering. In a study aimed at mitigating stereotypes, researchers employed post-processing techniques to modify portraits and enhance diversity [43]. However, it was observed that making elemental changes was also effective in breaking stereotypes. Compared to addressing a single character, the exchange of stereotypes between two characters proved to be more targeted and efficient.

In this third phase, the method is experimentally testable only with archetypal characters. This limitation exists because the approach involves extensive retention and exchange of stereotypes. Theoretically, the more stereotypes that are retained and exchanged, the more communicative and original the designs become. Archetypes are particularly suited for this approach because they inherently encompass numerous stereotypes.

Additionally, the use of AI to summarize stereotypes, as explored in the second phase, further inspired this phase. Directly querying AI for relevant information can save time, as AI-based creative assistance is now commonplace. However, the results generated by AI often reflect entrenched stereotypes. If not critically processed, the outcome may largely perpetuate existing stereotypes. Alternatively, AI can be used to reverse the process by summarizing and analyzing the information.

The third phase of modifying an existing design involves replacing stereotypes using methodological criteria. In this phase, the units of analysis are the stereotypical elements summarized through ChatGPT. The rationale for using ChatGPT-4 to summarize stereotypes in Phase 3 is that, in the previous phase, AI-based image generation was employed to identify stereotypical visual elements, while ChatGPT-4 was used to summarize these stereotypes in textual form for comparison.

Phase 3 selects which stereotypes to replace according to the following criteria:

1. Preference is given to stereotypes with significant visual differences within the same context. If the visual differences are minimal, replacement is unnecessary. For example, a tyrant and a mad scientist may both have scars—one from an experimental burn and the other from a war wound. This difference is not significant enough to warrant replacement.
2. Stereotypes that signify identity should not be replaced, as they help communicate the intended message of the design. For instance, a crown symbolizes the identity of a tyrant king, while elements like a palace or a laboratory indicate specific roles.
3. The greater the number of elements replaced, the more original the design. Following these criteria, replacing more elements results in a more unique design.

Next, stereotypes to retain are selected based on the following criteria:

1. If stereotypes meet Criterion 1 for replacement, they should not be retained.
2. Priority is given to stereotypes that reflect the character's identity. For example, a tyrant would retain stereotypes related to both madness and scientism, while a mad scientist would retain those evoking sadism and domination. A stereotype is considered emblematic of the character's identity if it alone can evoke recognition of that identity. The less reliance on combining it with other stereotypes, the more desirable it is.
3. Preference is given to stereotypes based on their frequency of occurrence. Stereotypes that occur more frequently are more effective in conveying the character's identity, as long as they meet Criterion 1 for retention.

To test the feasibility of this method, two prototypes were manually selected: the Tyrant and the Mad Scientist. Then, two additional archetypes were randomly selected using ChatGPT, and another round of testing was conducted.

We further elaborate on these benchmarks using the examples of the "tyrant" and the "mad scientist." First, according to the exchange criteria, stereotypes to be swapped should be selected based on the following guidelines: stereotypes that share the same type or context but exhibit significant visual differences should be prioritized, while those directly tied to identity are excluded. The results of the screening process are as follows: Physique: The tyrant's "tall and strong body" was exchanged for the mad scientist's "exaggerated body." Eyes: The tyrant's "cold eyes" were swapped for the mad scientist's "crazy eyes." Posture: The tyrant's "severe posture" was replaced with the mad scientist's "strange posture." Accessories: The tyrant's "metallic accessories" were exchanged for the mad scientist's "glasses or goggles." Appearance: The tyrant's "dignified appearance” was swapped for the mad scientist’s “ragged appearance” and “deranged smile.”

Next, stereotypes to be retained were chosen based on their relevance to character identity: Tyrant: Stereotypes emphasizing the tyrant’s status and brutality were prioritized. These include “luxurious robes and uniforms” and “ornate thrones and palaces,” which symbolize power and authority. Additional elements such as “armed” and “owning slaves and servants” further highlight the tyrant’s violent and oppressive nature.

Mad Scientist: Stereotypes reflecting the mad scientist’s identity were prioritized. Common elements such as the “white coat,” “chaotic lab environment,” and “luminescent or discharging devices” strongly convey the essence of the character. Although “wearing glasses or goggles” frequently appears, it does not fully represent the scientist’s identity. Thus, “crazy hair,” which occurs in 85% of cases, was included as a complementary stereotype. Based on the above criteria, character designs were developed. Based on the principles of retaining or altering stereotypes, we have developed a set of neutral prompt statements. Examples of these prompts are provided below: Tyrant: “A man in luxurious royal robes and uniform sits in a palace, seated on an ornate throne. He holds a sword, has a lean body, wears glasses, and has crazy eyes and a deranged smile. Behind him, people follow.” Mad Scientist: “In a laboratory cluttered with cluttered piles of luminescent and discharge devices and literature, a muscular man wears a lab coat with metal accessories, cold eyes and a crazy haircut.”

In the second and third phases of the study, neutral language was used when designing the prompts, and an objective perspective was explicitly sought to avoid reinforcing stereotypes due to biased training data in AI models. Additionally, emotional or subjective tones were minimized to prevent inadvertently reflecting stereotypes. A more specific documentation of the methodology of this phase can be found in our previous paper [44].

A. Pre-experimentation

To ensure adequate statistical power and reliability, a preliminary small-scale study was conducted to estimate the required sample size for the experiment. The procedure followed is outlined below:

Initially, a small-scale experiment was carried out, where questionnaires were distributed to 30 subjects. Of these, 25 subjects successfully recognized the characters, which provided an estimate of the expected recognition rate for the study.

Expected Recognition Rate (p): 0.80

Threshold Percentage for Hypothesis Testing (p_0): 0.50

Significance Level (α): 0.05

Statistical Power ($1-\beta$): 0.80

Using these parameters, we calculated the required sample size to ensure the experiment possessed sufficient statistical power. The sample size calculation was performed using Python. The following code was used for the calculation:

```
from scipy.stats import norm
import math

p = 0.80
p0 = 0.50
alpha = 0.05
power = 0.80

z_alpha = norm.ppf(1 – alpha / 2)
z_beta = norm.ppf(power)

n = ((z_alpha * math.sqrt(2 * p0 * (1 – p0)) + z_beta * math.sqrt(p * (1 – p) + p0 * (1 – p0))) ** 2) / ((p – p0) ** 2)
n = math.ceil(n)
n
```

The responses obtained were $n = 42$. This means that at least 42 questionnaires were needed to ensure that the study had sufficient statistical power and reliability to validate the success of the methodology.

B. Questionnaire

The questionnaire consists of two 7-point Likert scale questions and two open-ended questions. Initially, the examinee is shown the work and asked to identify the character(s) depicted (or more than one). The test taker is then informed of the correct answer, which activates stereotypes associated with the character. Devine concluded from several experiments that stereotypes are automatically activated by symbolic equivalents. For example, certain words can easily trigger stereotypes about Black people [45]. Since stereotypes are a form of knowledge, there is no inherent contradiction between stereotypes and the beliefs individuals hold. In other words, showing subjects the name of a character can activate the associated stereotype in their minds. Patriciaz [46] notes that even if an individual perceives a stereotype that does not align with their personal beliefs, the stereotype may still exist as a form of knowledge without necessarily leading to prejudice. Patriciaz [46] emphasizes that stereotypes, even when they do not match one’s personal beliefs, persist in the creator’s work. This insight helps explain the results of media literacy education experiments, where students were able to identify and criticize stereotypes in others’ work, but continued to incorporate them into their own creative outputs. This suggests that the ability to recognize a character is a key condition for studying the deconstruction of character stereotypes.

Next, the participants were asked to respond to the following questions:

1. How certain are you in recognizing the character?

The responses to this question help determine whether the method effectively facilitates clear and reliable communication of the work.

☐ Completely identifiable
☐ Largely identifiable
☐ Not sure
☐ Somewhat identifiable
☐ Not identifiable

2. Do you think the character differs from your stereotypical image?

This question assesses whether the method preserves the originality of the work.

☐ Completely changed

☐ Substantially changed

☐ Not sure

☐ Has not changed much

☐ Remains largely unchanged

3. Please share your thoughts about this experiment.

This final open-ended question is intended to explore the participants' subjective perspectives on the experiment in more detail. The answers to the first two questions help determine whether the method ensures the communicability of the work, while the answer to the fourth question indicates whether it guarantees the originality of the work.

The purpose of the survey was to evaluate whether the methods applied at each Phase were effective in reducing stereotypes. Two key criteria were used for evaluation:

1. Communicability: This criterion assessed whether the audience successfully recognized the character after the application of the method.

2. Originality: This criterion evaluated whether the audience observed any change in the character after the method was applied and whether the character deviated from the stereotypical image.

The results of the questionnaire were analyzed from both objective and subjective perspectives, focusing on two key aspects: communicability and originality in the design. Why are communicability and originality essential criteria for success in breaking stereotypes? This study draws on semiotics, communicative functionalism, and cognitive psychology. From these perspectives, Hauser regards stereotypes as visual "formal principles" (stehende Formeln), or rules for image creation. From a communication standpoint, the flow between the artist's creation and the viewer's interpretation relies on both parties sharing and understanding a common visual representation schema. Hauser likens this to a "language of art" [47]. The "language of art" is mutually understood by both the artist and the viewer, but the cost of this smooth communication is the loss of some originality in the work.

Stereotypes, on the other hand, are standardized visual images and designs, fixed forms, and recurring patterns. Gombrich describes stereotypes as "patterns" or "formulas," which are stable structures in visual representation and artistic expression. This comparison of stereotypes to formulas inspired the current study. In other words, if all the stereotypical elements of a character were changed arbitrarily, the character's identity would become difficult to recognize. Therefore, to successfully break a stereotype, some elements of the stereotype must remain so that the character's identity can still be understood. Otherwise, the result would simply be a new character inspired by the original. From Hauser's perspective, this scenario highlights a conflict in which originality might overshadow communicability. If communication is overly prioritized, the character design will remain formulaic and lack originality [48]. Thus, artists must carefully navigate the balance between originality and communicability in the creative process.

- Communicability:

Objective Evaluation: Question 1 of the questionnaire—blind test. Participants are asked to identify characters, and their recognition rate is evaluated. Evaluation Criteria: In usability studies related to visual design, a recognition rate of 70% is typically considered the benchmark for clarity and effectiveness of design elements [59]. Statistically, a 70% "majority threshold" is commonly regarded as indicative of design features being understood by most users. Furthermore, numerous industry guidelines recognize 70% as an acceptable benchmark for clear and intuitive design [50]. For instance, Wiedenbeck's research demonstrates that recognition rates exceeding 70% are typically viewed as a reliable indicator of an icon's intuitive design [51]. Therefore, a recognition rate of 70% or higher indicates that the design passes the objective communicability evaluation.

Subjective Evaluation: Question 2—If the participant successfully identified the character in Question 1, they are asked: "How confident are you in your identification of that archetype?" Survey Question 4 is an open-ended question: "Please share your thoughts about this design." Responses to Question 4 will be analyzed using Term Frequency (TF) analysis and sentiment lexicon methods. Evaluation Criteria: An objective recognition rate of at least 70%, followed by a subjective assessment of respondents' confidence in their identification.

Open-ended Feedback: Survey Question 4 asked participants, "Please share your thoughts on this design." Responses to this open-ended question will be analyzed using a Term Frequency (TF) approach and a sentiment lexicon to gain insights into their perceptions. An affective score greater than 0 indicates a positive emotional response toward the design.

- Originality:

Objective Evaluation: A controlled trial was conducted to assess the differences in design before and after applying the method. Evaluation Criteria: Objective judgment of the differences in design elements.

Subjective Evaluation: Survey Question 3 and other survey questions. Responses will be analyzed using text analysis techniques. Evaluation Criteria: Differences in design that can be visually judged, based on the subjectivity of the respondents.

C. Text Analysis

After the questionnaires were collected, responses to Question 4 were analyzed using an approach based on Term Frequency (TF) and the sentiment lexicon.

```
import jieba
from collections import Counter
import re

def remove_punctuation(text):
    text = re.sub(r'[^\w\s]', ", text)
    return text

def calculate_word_frequency(text, stop_words):
    text = remove_punctuation(text)
    words = jieba.lcut(text)
    words = [word for word in words if word not in stop_words and
len(word) > 1]
    word_counts = Counter(words)
    return word_counts

if __name__ == "__main__":
    chinese_text = " text "

    word_frequency = calculate_word_frequency(chinese_text,
stop_words)

print("Word Frequency Statistics:")
for word, count in word_frequency.items():
    print(f"Word '{word}' appeared {count} times.")
```

Based on the word frequencies already counted, an analysis was conducted using a sentiment lexicon-based approach. It is divided into the following steps

1. Load the sentiment lexicon: Using the already counted word frequencies, prepare a positive and negative sentiment lexicon containing words and their corresponding sentiment weights.

Continuing from the previous example, the positive words for communicability are summarized from the word frequencies: “aware”: 2, “fit”: 2, “associative”: 2, “immediately”: 2, “immediately”: 2. The negative words for communicability are: “hesitate”: –2, “worried”: –2, “Loch Ness”: –2 and “can’t see”: –2; “don’t know”: –2.

Positive words for originality can be summarized by word frequency: “novel”: 2, “innovative”: 2, “novelty”: 2, “creative”: 2, “break”: 2. Negative words for originality can be summarized as: “fit”: –2, “unchanged”: –2, “not broken”: –2 the same”: –2, “stale”: –2}.

2. Word matching and score calculation: Match the word frequency statistics with the sentiment dictionary and calculate the sentiment score of the text according to the matching results.

3. Sentiment score analysis: Determine the overall sentiment trend of the text by the value of the sentiment score. Scores greater than 0 are positive, scores less than 0 are negative, and scores equal to 0 are neutral. The results are used to determine the subject’s subjective sentiment tendency.

The code is as follows:

```
import jieba
from collections import Counter
import re

# Example sentiment dictionary
positive_words = {"realize": 2, "satisfies": 2, "associate": 2,
"immediately": 2, "quickly": 2}
negative_words = {"hesitate": -2, "conflicted": -2, "Loch Ness": -2,
"didn't notice": -2, "don't know": -2}
```

```
def remove_punctuation(text):
    text = re.sub(r'[^\w\s]', ", text)
    return text

def calculate_word_frequency(text, stop_words):
    text = remove_punctuation(text)
    words = jieba.lcut(text)
    words = [word for word in words if word not in stop_words and
len(word) > 1]
    word_counts = Counter(words)
    return word_counts

def sentiment_analysis(word_counts, positive_words, negative_words):
    sentiment_score = 0
    for word, count in word_counts.items():
        if word in positive_words:
            sentiment_score += positive_words[word] * count
        elif word in negative_words:
            sentiment_score += negative_words[word] * count
    return sentiment_score

if __name__ == "__main__":
    # Example Chinese text
    chinese_text = " text "
    # Common pronouns and stop words
    stop_words = {"I", "you", "he", "she", "it", "we", "you", "they",
"them", "those", "a", "an", "the", "is", "in", "there", "has", "and",
"also", "this", "that"}

    # Calculate word frequency
    word_frequency = calculate_word_frequency(chinese_text,
stop_words)

    # Print word frequency statistics
    print("Word Frequency Statistics:")
    for word, count in word_frequency.items():
        print(f"Word '{word}' appeared {count} times.")

    # Sentiment analysis
    sentiment_score = sentiment_analysis(word_frequency,
positive_words, negative_words)

    # Print sentiment score
    print(f"Sentiment Score: {sentiment_score}")

    # Analyze overall sentiment
    if sentiment_score > 0:
        print("Overall Sentiment: Positive")
    elif sentiment_score < 0:
        print("Overall Sentiment: Negative")
    else:
        print("Overall Sentiment: Neutral")
```

In this example, the AFFECT score is 6, indicating that the overall affective tendency of this sentence is positive from a communicative perspective. From this example, the subject’s subjective affective tendency was judged in terms of communicability. The same goes for originality.

IV. Result and Discussion

After the proposed design was completed, 42 anonymous subjects were asked to participate in an online questionnaire for this Phase. Table I below summarizes the data obtained from the experiment.

The survey results showed that all three methods ensured communication and originality of the work. This proves the feasibility and effectiveness of the three methods. A comparison of the data for Q123 is shown below in Fig. 1.

TABLE I. SUMMARY OF EXPERIMENTAL DATA

	Questionnaire	D	W	T	M S	E	AI A
Q1	people who recognized character in the first response	42	42	25	17	32	7
	people who recognized character in the second response	/	/	14	13	7	17
	Total number of people who correctly identified character	42	42	39	30	39	24
Q2	people who fully identified character	42	42	11	6	38	7
	people who basically recognized character	/	/	28	24	1	17
Q3	people who think character has completely changed	15	31	38	21	37	18
	people who think that character has basically changed	27	11	4	13	5	14
	Total number of people who think character has changed	42	42	42	34	42	32

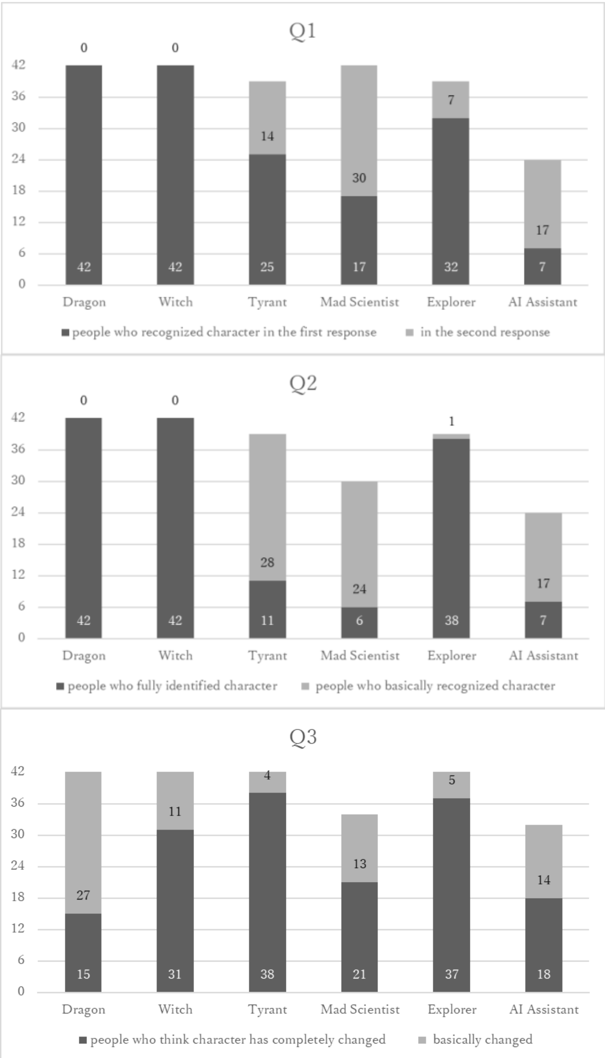

Fig. 1. A comparison of the data for Q123

The data from Q1 and Q2 reflect the communication skills associated with the three methods. As shown in the chart, the Phase 1 and Phase 2 methods demonstrate superior communicability. The data from Q3 assess the originality of these approaches, with the Phase 3 method excelling in this area. The Phase 1 and Phase 2 methods also show a degree of success. An analysis of the responses to Q4 indicates that all three methods are effective overall. Regarding communicability, subjects' positive emotional responses are ranked as follows: Phase 1 > Phase 2 > Phase 3. Conversely, in terms of originality, the ranking is Phase 3 > Phase 2 > Phase 1. These results further support the theory that communicability and originality are negatively correlated.

The subjective and objective evaluations for each phase are listed below:

- Phase 1: Dragon

Communicability

Objective Evaluation: The responses to the first questionnaire question demonstrated that the communicability of the methodology at this phase was highly effective, with a 100% recognition rate.

Subjective Evaluation: The responses to the second questionnaire question further confirmed the success of the methodology in terms of communicability. All participants (100%) reported feeling confident in their ability to fully recognize the characters.

Text Analysis: Responses to Question 4 were analyzed using a Term Frequency (TF) approach and sentiment lexicon analysis. The resulting sentiment score was 44, indicating an overall positive affective trend regarding communicability at this phase.

Originality

Objective Evaluation: The control trial revealed that character stereotypes differed significantly from the design after applying the methodology, indicating an impact on the originality of the design.

Subjective Evaluation: Responses to the third questionnaire question indicated that all participants (100%) believed that the character's stereotype had changed following the use of the methodology. Among them, 35.71% believed the character had completely changed from the stereotype, while 64.28% felt the character had substantially changed.

Text Analysis: Responses to Question 4 were analyzed using a TF approach and sentiment lexicon analysis. The effect score was 20, suggesting an overall positive sentiment trend regarding the originality of the design at this phase.

- Phase 2: Witch

Communicability

Objective Evaluation: Responses to the first question of the questionnaire demonstrated that the methodology's communicability at this phase was highly effective, achieving a 100% recognition rate.

Subjective Evaluation: The responses to the second question further confirmed the strong communicability of the methodology, with all participants (100%) reporting that they fully recognized the characters.

Text Analysis: Responses to Question 4 were analyzed using a TF approach and sentiment lexicon analysis. The resulting sentiment score was 40, indicating a positive affective trend in terms of communicability at this phase.

Originality

Objective Evaluation: The controlled trial revealed that character stereotypes differed significantly from the design after applying the methodology, suggesting that the method effectively altered the original stereotypes.

Subjective Evaluation: The responses to the third question indicated that all participants (100%) perceived the character's stereotype as being different from the design after the methodology was applied. Of these, 73.81% believed that the character had completely changed compared to the stereotype, while 26.19% believed that the character had substantially changed.

Text Analysis: Responses to Question 4 were also analyzed using a TF approach and sentiment lexicon analysis. The sentiment score of 34 suggests a positive affective trend in terms of originality at this phase.

- Phase 3: Tyrant

Communicability

Objective Evaluation: The responses to the first question of the questionnaire indicated that the communicability of the methodology at this phase was highly effective, with a recognition rate of 92.85%.

Subjective Evaluation: The responses to the second question suggested that the methodology's communicability remained strong. While 26.19% of the participants felt they fully recognized the characters, 66.66% believed they had basically recognized them.

Text Analysis: Responses to Question 4 were analyzed using a TF approach and sentiment lexicon analysis. The resulting sentiment score was 34, indicating an overall positive affective trend in terms of communicability at this phase.

Originality

Objective Evaluation: The controlled trial revealed that the character stereotypes had changed significantly after the application of the methodology, demonstrating its impact on the originality of the design.

Subjective Evaluation: Responses to the third question showed that 100% of the participants believed that the stereotype of the character differed from the design after using the methodology. Of these, 90.47% felt the character had completely changed compared to the stereotype, while 9.52% thought the character had basically changed.

Text Analysis: The responses to Question 4 were also analyzed using the TF approach and sentiment lexicon analysis. The sentiment score was 42, indicating a strong positive trend in originality at this phase.

- Phase 3: Mad Scientist

Communicability

Objective Evaluation: The responses to the first question of the questionnaire showed that the methodology's communicability at this phase was effective, with a recognition rate of 71.42%.

Subjective Evaluation: The responses to the second question revealed that 14.28% of the participants felt they fully recognized the characters, while 57.14% believed they had basically recognized them.

Text Analysis: The responses to Question 4 were analyzed using a TF approach and sentiment lexicon analysis. The resulting sentiment score was 16, suggesting a positive, albeit less robust, affective trend in terms of communicability at this phase.

Originality

Objective Evaluation: The controlled trial demonstrated that character stereotypes had changed following the use of the methodology, supporting the method's effectiveness in enhancing originality.

Subjective Evaluation: According to the responses to the third question, 80.95% of participants thought the character's stereotype had changed after the methodology was applied. Of these, 14.28% felt the character had completely changed, while 57.14% believed the character had basically changed.

Text Analysis: The responses to Question 4 were analyzed using the TF approach and sentiment lexicon analysis. The sentiment score was 26, reflecting a positive trend in originality, albeit weaker compared to the "Tyrant" character.

- Phase 3: Explorer

Communicability

Objective Evaluation: Responses to the first question of the questionnaire indicate that the communicability of the methodology at this phase was excellent, with a recognition rate of 92.85%.

Subjective Evaluation: The second question further confirms the methodology's strong communicability, as 90.47% of the participants felt they fully recognized the characters.

Text Analysis: Responses to Question 4 were analyzed using a TF approach and sentiment lexicon analysis. The sentiment score of 38 indicates a positive affective trend in terms of communicability at this phase.

Originality

Objective Evaluation: The control trial demonstrated that character stereotypes differed significantly from the design after the application of the methodology, supporting the effectiveness of the method in enhancing originality.

Subjective Evaluation: The third question revealed that 100% of participants believed the stereotype of the character differed from the design after using the methodology. Of these, 88.09% thought the character had completely changed compared to the stereotype, while 11.90% believed the character had basically changed.

Text Analysis: The responses to Question 4 were analyzed using TF and sentiment lexicon analysis. The sentiment score of 40 indicates a strong positive trend in originality at this phase.

- Phase 3: AI Assistant

Communicability

Objective Evaluation: Responses to the first question of the questionnaire indicate that the communicability of the methodology at this phase was moderate, with a recognition rate of 57.14%.

Subjective Evaluation: According to the second question, 40.47% of participants felt they fully recognized the characters, while 16.66% felt they basically recognized them. This suggests that the methodology had moderate effectiveness in terms of communicability.

Text Analysis: The responses to Question 4 were analyzed using a TF approach and sentiment lexicon analysis. The sentiment score of 10 indicates a positive but modest affective trend for communicability at this phase.

Originality

Objective Evaluation: The controlled trial confirmed that character stereotypes had changed following the application of the methodology, indicating an impact on originality.

Subjective Evaluation: The third question revealed that 76.19% of participants believed the stereotype of the character differed from the design after using the methodology. Of these, 42.85% felt that the character had completely changed, while 33.33% believed the character had basically changed compared to the stereotype.

Text Analysis: Responses to Question 4 were analyzed using TF and sentiment lexicon analysis. With a sentiment score of 14, the overall affective trend in terms of originality is positive but relatively mild.

V. Conclusion

This study begins by addressing the impact of stereotyping. Numerous studies have shown that media stereotyping has several negative effects on those who are stereotyped. Both Social Identity Theory and Initiation Theory, as part of Media Effects Theory, suggest that these negative impacts contribute to prejudice and influence behavior. Beyond media stereotypes, the rapid development of Artificial Intelligence (AI) in recent years has also played a significant role in reinforcing stereotypes. Many AI image generators have become widely used across creative and technical industries. These tools offer new possibilities, allowing individuals to generate visually striking images without the need for artistic expertise. However, images generated from large datasets often propagate additional stereotypes. Furthermore, AI's integration into creative activities has become commonplace, with educators encouraging students to draw inspiration from AI-generated images. Yet, the results of these image generation processes often reflect numerous stereotypes. Without critical and creative thinking, much of the resulting work remains stereotypical, repetitive, and formulaic.

Against this backdrop, this study emphasizes the necessity of developing specific methods to subvert stereotypes. A major contributing factor to the persistence of stereotypes and prejudices is a superficial understanding or lack of knowledge about others. Therefore, this study investigates specific approaches to overturning stereotypes. It is posited that breaking stereotypes requires a thorough examination of relevant knowledge about the subject. Since such investigations are often time-consuming, this study proposes AI-based methods to expedite the process. Additionally, the advantages and disadvantages of these methods are discussed, alongside considerations of which methods are most appropriate for different contexts.

In general, this study explores three methods for deconstructing visual stereotypes across three phases, with the data compared and summarized. The methods include: 1. Analyzing textual media to deconstruct stereotypes of characters whose origins conflict with current representations; 2. Examining multiple media sources and historical records to address stereotypes of characters that have evolved over time; 3. Using ChatGPT-4 to summarize archetypal character stereotypes, followed by a deconstruction to assess whether these stereotypes are maintained or replaced.

This study aimed to achieve three primary objectives. First, it proposed specific methods to reduce standardized visual elements—or stereotypes—in current designs. These methods generated new visual representations of characters while ensuring that they remained recognizable to the audience. Second, the study sought to mitigate the negative effects associated with stereotypes. Experimental data showed that the characters used in the experiment deviated from existing stereotypes. Since it takes time for elements of current stereotypes to shift and be accepted as new norms, it is theoretically possible to reduce the negative impacts of existing stereotypes by gradually incorporating the newly developed character concepts into creative work. The third objective was to analyze and compare the experimental data, summarizing the strengths and weaknesses of the methodologies, as well as the contexts in which they are most appropriately applied.

Through theoretical analysis and experimentation, this study draws the following conclusions:

1. Based on the results of the experiment involving the "Dragon" in the first stage, it is evident that stereotypes about a character can be altered by providing more information through written media.

2. From the results of the experiment involving the "Witch" in the second stage, it is clear that a character's stereotype can be modified by gaining additional information through various media and by leveraging the character's existing stereotype.

3. Based on the experimental results from the third stage, which focused on the "Tyrant" and four other archetypes, stereotypes can be effectively deconstructed by summarizing, exchanging, and retaining the stereotypes of the two archetypes through AI.

In conclusion, this study confirms that obtaining more information about characters is crucial for breaking stereotypes. More information, in this context, refers to any details about any aspect of a character recorded in written media, as long as it contributes to a deeper understanding of that character. Furthermore, we proposed a method to assess the success or failure of stereotype-busting based on communicability and originality. The objective of breaking stereotypes in character design was achieved through the three methods tested.

The results of the experiment demonstrated that all three methods were effective in breaking down stereotypes. Below are the advantages, disadvantages, and recommended applications for each method:

- Phase 1

The method used for the “Dragon” involved reviewing textual media to gather more information about dragons and using this information to challenge current stereotypes by restoring the original visual representation.

Pros: Excellent communication and originality.

Cons: Time-consuming.

Best suited for: Design projects where the current stereotype does not align with the original, but it can also be applied in other scenarios. Since extensive data research is crucial for breaking down stereotypes, this method is very effective when ample design time is available.

- Phase 2

The “Witch” method utilized AI-generated images to summarize existing witch stereotypes in character design. These stereotypes served as clues to examine the history of witch’s stereotypes, using past stereotypes to challenge current ones.

Pros: Retains many stereotypes while incorporating past ones, making this the most communicative of the three methods. Originality ranks second.

Cons: Although less time-consuming than the first method, it still requires significant time due to the targeted investigation and exchange of stereotype histories.

Best suited for: Design subjects with long histories and evolving stereotypes.

- Phase 3

This method employed ChatGPT-4 to collect stereotypes of the objects and analyze their frequency in visual representations. Based on frequency and a set of criteria, stereotypes were either retained or replaced to break down each stereotype.

Pros: The least time-consuming method. The exchange of two characters stereotypes for each other makes this approach efficient. Originality is preserved.

Cons: Overall, the communication and originality of this method are slightly inferior to the previous two. Currently, it is limited to prototypical characters.

Best suited for: Design projects involving prototypical or stereotypical characters. This method is recommended when time is short.

In this study, the effectiveness of three stereotype-breaking methods was tested. While the third method addressed the issue of time consumption, it still presents some limitations. Although the first two methods require substantial time investment, both theoretical analysis and experimental data indicate that they are the most effective in ensuring the originality of the work. These methods are particularly suitable for large projects where a solid foundation and extended design time are essential.

To save time in the third stage, ChatGPT-4 was used to summarize stereotypes and their frequency of occurrence. However, compared to AI image generation, we observed that some prominent and frequently occurring stereotypes were not captured by ChatGPT-4. For example, the stereotypes of Tyrant and Scientist were missing key characteristics such as being male, middle-aged, or older—traits that are readily apparent in AI-generated images and occur frequently. Additionally, AI image generation proved more challenging and time-consuming when extracting 42 stereotypes compared to ChatGPT-4. Therefore, future studies will consider combining ChatGPT-4 and AI image generation to more comprehensively summarize character stereotypes.

This study primarily focuses on breaking down visual stereotypes. However, stereotypes in media manifest in various forms, such as characterization and plot development. Future research will investigate these stereotypes and explore methods to deconstruct them. These stereotypes are more complex than visual stereotypes, particularly because characterization and plot development are closely intertwined and difficult to analyze separately. Nevertheless, this study provides a foundation for such research, including the criteria of originality and communication for stereotype-busting and the use of AI. By integrating insights from other disciplines, such as narratology, it is believed that effective methods for breaking down these stereotypes can ultimately be developed.

Our study revealed that using ChatGPT-4 alone to summarize role stereotypes often results in the omission of more subtle or less obvious stereotypes, such as gender. In contrast, AI image generation provides a more detailed enumeration of stereotypes, along with an approximate probability of their occurrence. Therefore, combining both methods enables each to compensate for the other’s shortcomings in stereotype summarization, leading to a more comprehensive stereotype mapping.

The first and most challenging aspect of applying these methods to non-visual stereotypes lies in defining the relevant stereotype units in relation to the plot. For example, in Phase 3, a method of exchanging and retaining stereotypes was applied. If this approach were extended to deconstructing narrative stereotypes, a key question arises: what constitutes the unit to be exchanged? Is it a plot twist, the background setting, or something else? For non-visual stereotype mitigation, we propose initially summarizing the “plot” of a particular type of media. For example, this could involve outlining the stereotypical plots of horror novels or romantic movies. This step can be efficiently completed using tools such as ChatGPT-4 and AI-based image generation. Following this, strategies to deconstruct these stereotypical plots are still under exploration. We argue that deconstructing narrative and plot stereotypes presents a more complex challenge.

In general, this study addresses the limitations and inertia of traditional design thinking by promoting a character-based approach that encourages critical and dialectical thinking among creators. It also provides valuable reference points and inspiration for future research on stereotypes.

Prospects and Limitations of AI Technology Application in This Field: Despite the growing body of research on AI-induced stereotyping, completely eliminating stereotypes in AI remains a challenging task.

This is particularly evident in creative writing, where the limitations of AI become apparent, especially in avoiding stereotypical outcomes. In this study, rather than attempting to eliminate stereotypes, we leveraged the stereotypes embedded in AI. AI was employed to quickly summarize current character stereotypes, saving significant time in the research process. This opens up two possible future research directions:

1. Exploring methods to directly break down stereotypes using AI with minimal human judgment.
2. Addressing the communicability issue identified in the third phase of our method, as its adaptability was significantly lower compared to the first two phases.

Additionally, ensuring the generalizability and scalability of our approach for future development is a crucial consideration, particularly since the current method involves manual decision-making. We plan to collaborate with the broader research community to gather feedback and refine our methodology. By incorporating insights from a diverse group of researchers and practitioners, we aim to enhance the reliability and scalability of this approach.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

XF conducted the research, analyzed the data and wrote the paper. MM provided suggestions for academic revisions and validated the final version of the paper. All authors had approved the final version.

REFERENCES

[1] L. Marinucci, C. Mazzuca, and A. Gangemi, "Exposing implicit biases and stereotypes in human and artificial intelligence: State of the art and challenges with a focus on gender," *AI & Soc.*, vol. 38, pp. 747–761, 2023. https://doi.org/10.1007/s00146-022-01474-3

[2] A. Caliskan, J. J. Bryson, and A. Narayanan, "Semantics derived automatically from language corpora contain human-like biases," *Science*, vol. 356, no. 6334, pp. 183–186, 2017.

[3] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proc. Conference on Fairness, Accountability and Transparency*, 2018.

[4] D. DeSilver. (2023). What the data says about food stamps in the U.S. [Online]. Available: https://www.pewresearch.org/short-reads/2023/07/19/what-the-data-says-about-food-stamps-in-the-u-s/?itid=lk_inline_enhanced-template#:~:text=Non%2DHispanic%20White%20people%20accounted,and%2035.8%25%20of%20child%20recipients

[5] A. Birhane *et al*., "Multimodal datasets: misogyny, pornography, and malignant stereotypes," arXiv preprint, arXiv:2110.01963, 2021.

[6] N. Tiku, K. Schaul, and S. Y. Chen. (2023). These fake images reveal how AI amplifies our worst stereotypes. [Online]. Available: https://www.washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/

[7] S. Ramasubramanian, A. Winfield, and E. Riewestahl, "Positive stereotypes and counter-stereotypes: Examining their effects on prejudice reduction and favorable intergroup relations," in *Media Stereotypes: From Ageism to Xenophobia*, 2020, pp. 258–276.

[8] T. L. Dixon, "Crime news and racialized beliefs: Understanding the relationship between local news viewing and perceptions of African Americans and crime," *Journal of Communication*, vol. 58, pp. 106–125, 2008.

[9] Y. Kawai, "Stereotyping Asian Americans: The dialectic of the model minority and the yellow peril," *Howard Journal of Communications*, vol. 16, no. 2, pp. 109–130, 2005.

[10] S. Ramasubramanian, "Television viewing, racial attitudes, and policy preferences: Exploring the role of social identity and intergroup emotions in influencing support for affirmative action," *Communication Monographs*, vol. 77, no. 1, pp. 102–120, 2010.

[11] S. Trimble, *The People: Indians of the American Southwest*, Santa Fe: School of American Research Press, 1993, p. 58.

[12] T. Robinson and C. Anderson, "Older characters in children's animated television programs: A content analysis of their portrayal," *Journal of Broadcasting & Electronic Media*, vol. 50, no. 2, pp. 287–304, 2006. doi: 10.1207/s15506878jobem5002_7

[13] S. Jörg and L. Schleussner, *Film and Stereotype: A Challenge for Cinema and Theory*, Columbia University Press, 2011, p. 13.

[14] R. Barthes, *S/Z: An Essay* (French 1970), trans. R. Miller (New York: Hill and Wang, 1974, p. 20.

[15] N. Signorielli, "Race and sex in prime time: A look at occupations and occupational prestige," *Mass Communication and Society*, vol. 12, no. 3, pp. 3327–352, July 2009.

[16] E. Behm-Morawitz and M. Ortiz, "Race, ethnicity, and the media," in *The Oxford Handbook of Media Psychology*, K. E. Dill Ed. 2013, pp. 252–266.

[17] G. Gerbner and L. Gross, "Living with television: The violence profile," *Journal of Communication*, vol. 26, no. 2, pp. 172–199, 1976.

[18] M. J. Lee, S. L. Bichard, M. S. Irey, H. M. Walt, and A. J. Carlson, "Television viewing and ethnic stereotypes: Do college students form stereotypical perceptions of ethnic groups as a result of heavy television consumption?" *Howard Journal of Communications*, vol. 20, no. 1, pp. 95–110, 2009. doi: 10.1080/10646170802665281

[19] E. Behm-Morawitz and D. Ta, "Cultivating virtual stereotypes? The impact of video game play on racial/ethnic stereotypes," *Howard Journal of Communications*, vol. 25, no. 1, pp. 1–15, 2014.

[20] O. Johnston and F. Thomas, *The Disney Villain*, New York: Hyperion Press, 1993.

[21] M. Li-Vollmer and M. E. LaPointe, "Gender transgression and villainy in animated film," *Popular Communication*, vol. 1, no. 2, pp. 89–109, 2003.

[22] Ranker. The greatest movie villains of all time. [Online]. Available: https://www.ranker.com/crowdranked-list/the-best-movie-villains-of-all-time

[23] G. Kyriakou, V. Drivelou, A. Glentis, "Villainous hair: Ba (l) d to the bone—Would they be so evil if they had hair?" *British Journal of Dermatology*, vol. 184, issue 1, pp. 156–157, 2021. https://doi.org/10.1111/bjd.19508

[24] K. Montgomery, C. White, and A. Thompson, "A mixed methods survey of social anxiety, anxiety, depression and wig use in alopecia," *BMJ Open*, vol. 7, e015468, 2017.

[25] K. E. Miller, "Children's behavior correlates with television viewing," *American Family Physician*, vol. 67, pp. 593–594, 2003.

[26] N. Signorielli, "Television's gender role images and contribution to stereotyping: Past, present, future," in *Handbook of Children and the Media*, D. G. Singer and J. L. Singer, Eds. Thousand Oaks, CA: Sage, 2001, pp. 341–358.

[27] N. Falchikov, "Youthful ideas about old age: An analysis of children's drawings," *International Journal of Aging and Human Development*, vol. 31, pp. 79–99, 1990.

[28] K. E. Miller, "Children's behavior correlates with television viewing," *American Family Physician*, vol. 67, pp.593–594, 2003.

[29] T. Robinson and C. Anderson, "Older characters in children's animated television programs: A content analysis of their portrayal," *Journal of Broadcasting & Electronic Media*, vol. 50, no. 2, pp. 287–304, 2006. doi: 10.1207/s15506878jobem5002_7

[30] M. Middlecamp and D. Gross, "Intergenerational daycare and preschoolers' attitudes about aging," *Educational Gerontology*, vol. 28, pp. 271–288, 2002.

[31] R. Brown, *Social Psychology*, London: Collier-Macmillan, 1965.

[32] P. F. Secord and C. W. Backman, *Social Psychology*, 2nd ed. Tokyo: McGraw-Hill, 1974.

[33] I. Ramos-Soler, C. López-Sánchez, and T. Torrecillas-Lacave, "Online risk perception in young people and its effects on digital

behaviour," *Comunicar*, vol. 26, no. 56, pp. 71–79, 2018. doi:10.3916/c56-2018-07 (in Spanish).
[34] J. Kahne and B. Bowyer, "Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation," *American Educational Research Journal*, vol. 54, no. 1, pp. 3–34, 2017.
[35] E. Schiappa, P. B. Gregg, and D. E. Hewes, "The parasocial contact hypothesis," *Communication Monographs*, vol. 72, no. 1, pp. 92–115, 2005. doi:10.1080/0363775052000342544
[36] E. Scharrer and S. Ramasubramanian, "Intervening in the media's influence on stereotypes of race and ethnicity: The role of media literacy education," *Journal of Social Issues*, vol. 71, no. 1, pp. 171–185, 2015. https://doi.org/10.1111/josi.12103
[37] A. Kavoori, "Media literacy, 'Thinking Television,' and African American communication," *Cultural Studies/ Critical Methodologies*, vol. 7, no. 4, pp. 460–483, 2007. doi: 10.1177/1532708607305038
[38] P. R. Hinton, *Stereotypes, Cognition and Culture*, 2000, p. 23.
[39] X. Feng and M. Murakami, "Design outside the stereotypes: Restoring the dragon from revelation," in *Proc. 2021 10th International Congress on Advanced Applied Informatics*, 2021. doi: 10.1109/IIAI-AAI53430.2021.00140
[40] Y. Wang, C. Wu, L. Herranz, J. van de Weijer, A. Gonzalez-Garcia, and B. Raducanu, "Transferring GANs: Generating images from limited data," in *Proc. the European Conference on Computer Vision (ECCV)*, 2018.
[41] D. Zowghi and F. da Rimini, "Diversity and inclusion in artificial intelligence," arXiv preprint, arXiv:2305.12728, 2023.
[42] X. Feng and M. Murakami, "Design that uses AI to overturn stereotypes: Make witches wicked again," *International Journal of Artificial Intelligence & Applications*, vol. 14, pp. 51–62, 2023. doi: 10.5121/ijaia.2023.14205
[43] X. Zhu, Z. Lei, and S. Z. Li, *Post-Processing of Face Swapping Generated by GANs*, 2018.
[44] X. Feng and M. Murakami, "Subverting characters stereotypes: exploring the role of AI in stereotype subversion," *International Journal of Artificial Intelligence & Applications*, vol. 14, pp. 13–23, 2023. doi: 10.5121/ijaia.2023.14502
[45] P. G. Devine and A. J. Elliot, "Are racial stereotypes really fading? The Princeton trilogy revisited," *Personality and Social Psychology Bulletin*, vol. 11, 1139, 1995.
[46] P. G. Devine, "Stereotypes and prejudice: Their automatic and controlled components," *Journal of Personality and Social Psychology*, vol. 56, no. 1, pp. 5–18, 1989. https://doi.org/10.1037/0022-3514.56.1.5
[47] A. Hauser, *The Sociology of Art* (German 1974), trans. Kenneth J. Northcott (Chicago: University of Chicago Press, 1982), pp. 29–31.
[48] E. H. Gombrich, *Art and Illusion: A Study in the Psychology of Pictorial Representation*, 1960, pp. 55–78.
[49] J. N. Kapferer, "The new strategic brand management: Advanced insights and strategic thinking," *Kogan Page*, 122, 2012.
[50] International Organization for Standardization, ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability, 1998.
[51] S. Wiedenbeck, "The use of icons and labels in an end user application program: An empirical study of learning and retention," *Behaviour & Information Technology*, vol. 18, no. 2, pp. 68–82, 1999. https://doi.org/10.1080/014492999119122