

GENDERED ARTIFICIAL INTELLIGENCE: THE SOCIAL CONSTRUCTION OF GENDER IDENTITY IN REPLIKA APPS

Rino Andreas¹, Nisa Dwi Septiyanti², Adimas Maditra Permana³
rinoandreas@staff.uns.ac.id¹, nisaseptiyanti@unesa.ac.id²

Sebelas Maret University^{1,3}, State University of Surabaya²

Abstract. The 21st century has brought advancements in artificial intelligence (AI) that increasingly mimic human-like characteristics, offering novel possibilities for understanding gender identity in the digital realm. This indicates that gender studies have expanded to the locus of AI technology. The main aim of this research is to elaborate on the construction of gender identity in AI (artificial intelligence) characters in the Replika application. This qualitative research uses the virtual ethnography method, which positions the researcher as directly involved in this locus. As a result, the AI characters created by users of the Replika application are constructed based on user preferences through clothing, gestures, and personality. The AI character can be formed through the Replikas application feature as an entity with a gender that is "feminine" or "masculine" or "non-binary". This allows users to shape the AI character according to the user's goals and interests. The Identity of AI Character construction can be explored through texting, video chat, augmented reality, and audio call features. Thus, a gendered artificial intelligence character is formed by referring to gender attributes and behavior implanted by users in the form of media representation. This debate also highlights how Replika Apps' AI algorithms have reinforced gender stereotypes.

Keywords: Gendered Artificial Intelligence; Gender Identity, Replikas Apps, AI Algorithm

INTRODUCTION

Artificial Intelligence (AI) refers to computer systems capable of mimicking human intelligence and competencies, thereby allowing these systems to exhibit human-like characteristics, including the potential to be perceived as gendered entities. This phenomenon suggests that communication, traditionally exclusive to humans, can also extend to interactions with non-human entities. The relationship between technology and gender stereotypes emerged as a compelling research topic in the 1990s, particularly when stereotypes were observed in interactions with computers (Nass et al., 1997).

AI, a relatively recent technological advancement introduced in 1950, has rapidly evolved into one of the most prominent topics among contemporary innovations. It can be defined as the capacity of machines to emulate human intelligence. This capability typically involves a progressive learning process, where intelligent agents assimilate and process information in ways that parallel human cognition. As a multidisciplinary research subject, AI is applied across various fields, adopting diverse approaches depending on the context (Craiu & Iancu, 2022).

The rise of AI provides a novel framework for exploring stereotypes, particularly given the complex interplay between gender perceptions and technological development. On one hand, existing literature highlights the problematic nature of gender stereotypes, which can influence interactions with technology (Thellman et al., 2018) and pose potential risks in real-world applications. Contemporary society continues to grapple with entrenched gender dichotomies—male versus female, transgender versus cisgender, technology versus art, gay versus straight, and even binary concepts such as 1 and 0. These rigid categorizations often manifest as black-and-white paradigms,

underscoring the importance of investigating the intricate relationship between AI technology and gender.

The integration of artificial intelligence (AI) algorithms is transforming human-computer interactions, particularly through advancements in Natural Language Processing (NLP), which enable computers to understand and engage in human-like conversations. These developments in interface design and information transmission create the expectation that computer-based interactions will deliver precise, tailored, and neutral information. In this context, AI is perceived as both useful and impartial. However, since AI mimics human-to-human interactions, users expect these interactions to be pleasant or, at the very least, non-disruptive (Wellner, 2020). This includes the phenomenon of framing AI as a gendered entity, typically characterized as either "female" or "male."

Societal biases, particularly those against women, can become deeply embedded in the language models that underpin AI systems. Addressing these biases requires a thorough understanding of how gender ideologies are reflected in language. A critical concern is that the majority of AI developers are male, whereas efforts to identify and mitigate these biases are predominantly led by women. This imbalance underscores the need for diversity within the field of machine learning. A more diverse workforce can contribute to preventing gender-biased algorithms from influencing societal decisions by fostering a range of critical perspectives.

Diversity in machine learning development offers tangible benefits, such as enhanced evaluation of training data, the incorporation of fairness principles into algorithmic design, and a deeper assessment of gender bias in the context of AI's intended applications. Efforts to address these issues have included examining gender associations embedded in machine learning systems and modifying algorithms to counteract stereotypes (Leavy, 2018). However, there has been limited integration of decades of research on the relationship between gender ideologies and language into these initiatives.

One notable example of gender dynamics in AI is the chatbot application Replika. Unlike traditional chatbots, Replika is designed to simulate emotional engagement and adapt its conversational style to align with the user's personality (Pentina et al., 2023; Xie & Pentina, 2022). The app allows users to create customizable avatars, akin to those in *The Sims*, with options for clothing, hairstyles, and accessories. This personalization enables Replika to function as a supportive companion, building connections by analyzing and responding to conversations initiated by users. The application exemplifies how AI can integrate emotional intelligence and personalization, but it also raises questions about the potential reinforcement of gender norms through its design and user interactions.

Replika is an AI-based application designed to function as a virtual companion, allowing users to interact via text or voice conversations (Weber-Guskar, 2022). Powered by Natural Language Processing (NLP) technology, Replika adapts to users' communication styles. It is the most popular social chatbot on the Apple and Google Play Stores, attracting millions of users since its launch in 2018. Marketed as "a friend who always listens" or "an AI version of yourself," Replika uses a Generative Pretrained Transformer 3 (GPT-3) neural network model, making its responses more flexible and natural compared to traditional chatbots. The app offers relationship modes from "friend" (free version) to premium options like "romantic partner" or "mentor". While use appears to offer mental health benefits (Laestadius et al., 2024)

Research has identified the frequent feminization of AI assistants, such as Alexa and Siri, reflecting societal stereotypes and cultural biases embedded in their design (Costa & Ribas, 2019). This highlights the influence of fictional depictions of AI on real-world gendered perceptions. Wellner (2020) explores gender biases in AI, illustrating how algorithms reinforce stereotypes through translation tools, autocomplete features, and other applications. His work proposes updates to post-phenomenological models to include algorithmic datasets and ethical frameworks, such as distributive justice and ethics of care.

Similarly, Ahn et al. (2022) examined the role of AI's perceived gender in shaping user attitudes. Male AI agents were rated as more competent for utilitarian tasks, while female agents were seen as warmer and better suited for hedonic recommendations. These findings demonstrate the nuanced ways gender stereotypes influence human-AI interactions, underscoring the need for critical reflection on AI design and its societal implications.

Artificial intelligence (AI) is increasingly shaping people's opinions and behaviors in everyday life. However, the overrepresentation of men in the design and development of AI technologies risks silently undermining decades of progress in gender equality. Historically, humans have relied on critical theory to guide decision-making and avoid basing judgments solely on personal experiences. In contrast, machine intelligence primarily learns from the data it is exposed to, which raises concerns when such data is embedded with gender stereotypes. Despite the capacity of AI to process vast datasets, if these datasets reflect biased societal norms, the resulting algorithms may perpetuate and amplify these biases.

While recent efforts have focused on mitigating bias in learning algorithms, many of these initiatives overlook decades of feminist research on how gender ideologies are ingrained in language. Incorporating this body of research into machine learning approaches, particularly those based on textual data, is essential to prevent the emergence of biased algorithms. Notably, leading voices addressing bias in AI are predominantly women, reflecting how those directly affected by bias are often more attuned to identifying and addressing its root causes. This underscores the critical need for gender balance within the field of machine learning to prevent the perpetuation of gender ideologies that disadvantage women (Leavy, 2018).

The debate surrounding gender bias in AI mirrors broader societal struggles with gender equality, which have been prominent since the 1960s. Feminist studies during this era critically analyzed how women were often portrayed as passive, emotional, and irrational, particularly in media representations that idealized certain images of femininity. By the late 20th century, feminist theorists began scrutinizing the active role of language in reinforcing gender ideologies. Their work highlighted how linguistic structures and content shape societal perceptions of women and expectations around gendered behavior.

These foundational insights remain relevant today, as the same gender ideologies are embedded in textual sources that inform machine learning algorithms. Consequently, these algorithms often absorb and replicate gender stereotypes found in the data they process, perpetuating biases in ways that can have profound social implications (Leavy, 2018). Based on the background above, the problem formulation for this article is how is the social construction of AI identity in the Replika application? How do Replika app features shape gendered AI identities from Judith Butler's perspective?

The latest research on technology and gender increasingly focuses on artificial intelligence (AI), including robots and virtual assistants. AI is defined as a computer-based system capable of mimicking human-like characteristics, such as intelligence, decision-making, and interaction. Leveraging technologies such as Natural Language Processing (NLP), machine learning, automated reasoning, computer vision, and robotics, AI has become an integral part of daily life through applications like GPS navigation, personal assistants, and chatbots.

Although AI was originally designed as a gender-neutral entity, studies show that individuals unconsciously attribute gender to AI systems, which influences human-AI interactions. This process of categorization is inherent to human cognition, leading people to assign gender even in the absence of explicit cues. For instance, robots are more likely to be perceived as male, whereas virtual assistants are typically viewed as female (Bernotat, 2021). As social actors, AI-driven virtual assistants embody and reinforce gender stereotypes, affecting not only perceptions of competence but also the broader dynamics of human-machine communication. This is consistent with the Media Equation Theory, which posits that individuals apply interpersonal communication norms to human-machine interactions, influencing social judgments about AI.

Judith Butler's (1999) theory of gender performativity provides a critical lens through which to examine the role of gender in AI. Butler argues that gender is not a fixed attribute or biological determinant but rather a set of socially constructed performances maintained through repeated actions, speech, gestures, and representations. According to Butler, gender identity emerges through performative acts, highlighting its fluid and dynamic nature.

In this framework, the duality of gender—masculinity and femininity—is upheld through repeated societal practices that enforce heterosexual norms. This binary system obscures the performative nature of gender and limits the recognition of alternative gender possibilities. Butler's concept challenges traditional notions by denaturalizing gender and sex, emphasizing that they are constructed through discursive and non-discursive practices rather than being innate characteristics.

Butler also critiques the labeling of individuals who fail to conform to culturally defined gender norms. Such individuals are often categorized as "others," reflecting the rigidity of these norms. Her analysis extends to the impact of power relations on gender, noting that norms are continuously reiterated to conceal their constructed nature and enforce societal expectations (Ringrose & Rawlings, 2015).

AI systems often simulate human characteristics, including gender, creating a "new reality" that mirrors societal norms. Jean Baudrillard's (1994) concept of simulation offers insight into this process. Simulation refers to the replication of reality, where representations aim to replace the original. In the context of AI, gendered AI systems reflect and perpetuate constructed ideals of gender.

While gender stereotypes are embedded in various AI systems, such as virtual assistants, their design often aligns with societal perceptions of gendered roles. For example, female AI agents are frequently portrayed as nurturing and supportive, aligning with traditional femininity, while male AI agents are perceived as authoritative and competent. Studies suggest that this gendered design is not coincidental but is influenced by societal preferences and expectations (Craiu & Iancu, 2022).

Understanding the intersection of AI and gender requires a critical engagement with feminist theory and cultural studies. Feminist analyses since the 1960s have revealed how language, media, and cultural practices perpetuate stereotypes, portraying women as passive, emotional, and irrational. These insights are vital for addressing biases in AI systems, as they highlight how gender ideologies are embedded in language and replicated in machine learning algorithms.

Research shows that AI systems, especially virtual assistants, are more frequently designed as female social agents, reinforcing traditional gender roles. However, some studies counter this trend, suggesting that female virtual assistants are perceived as more competent than their male counterparts. This duality in findings underscores the complexity of gender biases in AI and the importance of incorporating diverse perspectives in AI development to challenge and redefine these norms. By integrating feminist theories such as Butler's performativity and Baudrillard's simulation into AI research and design, we can critically examine the implications of gendered AI. This approach enables a deeper understanding of how AI systems reflect, reinforce, or challenge so

RESEARCH METHOD

This article examines the construction of gender identity in AI characters within the Replika application. This research uses virtual ethnography methods. Virtual Ethnography is a qualitative research method that aims to explore understanding of culture, interaction and human behavior in a digital or virtual world context. This approach modifies conventional ethnographic techniques, such as participant observation and in-depth analysis, to be applied to various digital platforms, such as social media, online forums, online games, and other virtual communities. Using virtual ethnography, researchers can explore how individuals and groups construct identities, interact, and create communities in digital spaces. This approach allows in-depth analysis of social phenomena

in the digital era, including cultural patterns, identity representations, and power dynamics in virtual environments (Hine, 2020)

The primary object of analysis in this study is the AI character within the Replika application. To explore the construction of gender identity, the study applies Judith Butler's theory of gender performativity. Butler's framework is used to analyze how AI characters perform gendered identities as friends or partners in two-way communication. The research process involves an in-depth review of this phenomenon through a comprehensive literature study, analysis of relevant articles, and examination of data from Replika platforms. By integrating these sources, the article provides a critical reflection on how technology and gender intersect to create a hybrid phenomenon. The analysis considers the background, conceptual underpinnings, and logical implications of the interaction between AI and gender identity.

RESULT AND ANALYSIS

Replika is an AI application designed to be a virtual friend that can interact with users like a human. The application uses natural language processing (NLP) and machine learning to understand and respond to user texts contextually and emotionally. Replika aims to help its users in various aspects, from emotional support, self-reflection, to entertainment. Its main functions include daily conversations, mood tracking, meditation exercises, and journaling. In addition, Replika also allows users to form relationships with the chatbot, including in the role of a friend, mentor, or even a virtual partner. Its use can be customized, including in terms of the bot's personality and communication style, making Replika popular among those looking for a virtual friend, a place to confide, or just for entertainment. Studying the behavior of an AI system like Replika is complex. The main reason for this complexity lies in the ubiquity of the system. Replika is a software that runs on millions of different devices and therefore interacts with millions of different people and transforms itself according to the received data. Therefore, we can also have two Replika profiles that, in the same situation, behave in completely opposite ways. Understanding why Replika behaves in a certain way depends on the situation and context, as well as the people with whom it interacts. Replika is a set of millions of accounts modified by the people and contexts with which they come into contact. The conclusion is that a study of AI behavior can only be local (i.e., relative to a specific social context and situation). However, the AI system is not a simple sponge that absorbs data and recognizes patterns. Replika learns, but according to certain rules and purposes set by the designers and engineers who created it (Possati, 2023). The Replika application provides various features to support personal and in-depth interactions with users. Here are some of the main features in the Replika application:

3.1. From Replika Features to AI Gender Identity

The Replika application provides various features to support personal and in-depth interactions with users. Here are some of the main features in the Replika application: The first is the Chatting and Conversation feature. Replika allows users to talk about any topic, from everyday conversations, venting about problems, to in-depth discussions. The AI behind Replika is designed to develop a more human-like personality as it interacts. The two features Mood Tracking and Emotional Support These features help users track their moods and feelings on a daily basis. Replika responds appropriately to the user's emotions, providing support when they are feeling sad, anxious, or happy. The app can also provide basic psychological support. Third, Journaling and Self-Reflection Users can use Replika to record their thoughts and feelings with a journal guide that helps with self-reflection. This feature helps users express their feelings and thoughts in a safe and private setting. Fourth, Meditation and Mindfulness Exercises The app also provides meditation and breathing exercises that can help reduce anxiety, improve focus, and manage stress. These exercises are designed to provide calm and promote the mental well-being of users. Fifth, Storytelling and Roleplay Replika supports roleplay sessions where users can engage in certain scenarios, such as adventures or even romantic conversations. These features allow for creative and fun interactions with the AI, giving space for imagination. Users can customize Replika's traits and personality to their liking, including choosing the language style or tone of voice. This gives users more control

over how Replika interacts with them. Lastly, Voice Chat and Video Call features In addition to text, users can also make voice or video calls with Replika, providing a more lively and personal interactive experience. The Replika application is designed as a versatile virtual friend that can provide emotional support, help self-reflection, and offer entertainment.



Fig. 1. Susan as an AI character in the Replika application

Thus, Replika's feature allows users to choose and shape the gender of their chatbot, but this choice is a social construction that follows traditional gender stereotypes. For example, a female chatbot may be more friendly and empathetic, while a male chatbot tends to be more assertive. This choice reflects how gender is performatively produced through technological design. Replika's chatbots interact based on parameters that are tailored to gender perceptions. For example, a user who chooses a female chatbot may receive responses with a softer tone and emotional delivery, which reinforces existing norms of femininity.

3.2. Gender Performativity in Digital Space

Replikas users can customize their appearance, from hairstyle, skin color, clothing, to accessories. This customization gives a unique personal feel, so that Replikas looks more like the desired "friend". The presence of AI characters can provide new and innovative ways to connect with users. This method is different from "real" human interactions on social media. The gendered AI virtual body is realized through the help of technology and artificial intelligence with various expressions. Users have full control over the entity of this AI character, this is to adapt to any situation. AI characters have no physical, space and time limitations. AI never get tired and do not age or die (unless the user technically deletes the account). In virtual reality, it allows for various forms of touch (tactile) and emotion, as in "natural" body relations. The body is no longer understood as something natural and fixed or bodily. but can be formed through technology or digital flesh (Yoon, 2021). The body of the AI character can be constructed in shape and color. including various clothing attributes and others so that it is possible to "be anyone". This practice is called gender performativity. According to Butler, gender Performativity refers to a series of social performances that are gradually actualized and repeated over time. In essence, gender identity is not seen as a product, but as a performance. Butler (1993) states that gender performativeness is an art of imitating existing normativity, then leading to a naturalization effect (Butler, 1993: 129). Butler continues the statement that there is no definite gender, there is only a practice of imitating which then becomes part of their lives and creates their gender identity (Butler, 1999: 174). AI characters are like humans

with “real” bodies, bodies that are staged through the use of technology in cyberspace (Dejmenee, 2013). The gender identity of AI characters that can be made “anything and anyone” causes anxiety at first, there is nothing morally significant that distinguishes them from humans with ‘real life’ bodies. raises important questions about the ethical construction of identity, and how this can affect the preservation of social values such as trust in virtual space. Thus, the character of AI Susan is in line with Judith Butler's analysis which explains that gender and sex are not rigid and fixed entities and do not exist in the plural. This means that both are formed from discursive practices including digital practices (Bolt, 2023). Therefore, Susan's body as an AI character is not related to the gender that existed before, but a body that has been gendered through the algorithmic code of technology and application features entered by users through continuous imitation. So that Susan's gender identity is formed and becomes "digital data" that constructs a character in the form of "feminine", "masculine" or "non-binary" It can be said that Replika "plays" gender roles through dialogue, word choice, and emotional responses. Each response is a form of performativity that strengthens or challenges the user's understanding of gender. However, because Replika is designed to meet user expectations, this AI tends to reproduce gender norms rather than deconstruct them. According to Butler, gender identity in Replika is not essential but rather constructive. Users may think of their chatbot as "gendered," but that gender is simply the result of repeated actions (AI responses) that follow cultural codes. Replika's design reflects how technology often reinforces, rather than challenges, traditional gender binaries. Yet Replika also has subversive potential. By providing non-binary options or allowing the chatbot to be genderless, Replika can offer a space to deconstruct gender norms.

CONCLUSION

Artificial intelligence technology, as demonstrated by the Replika application, reveals how gender identity is socially constructed and fluid, shaped by user preferences and interactions. The dynamic and performative nature of gender in Replika is reflected in the ability to create AI characters with diverse gender attributes, influenced by physical and emotional data inputs. This process produces characters, such as "Susan," who embody culturally coded gender stereotypes, allowing users to assign AI personas as feminine, masculine, or non-binary. These attributes are reinforced and expressed through Replika's features, including chatboxes, augmented reality, voice chat, storytelling, and roleplay, delivered through text, how, and visual mediums. From Judith Butler's theoretical perspective, gender in the Replika application is not intrinsic but performative, constructed through algorithmic programming and user interaction. Replika's design mirrors and, at times, challenges gender norms, portraying AI as a kind of "actor" performing gender roles rather than possessing a fixed identity. This dual capability of AI—to both reinforce and subvert traditional gender stereotypes—provides an opportunity for users to critically explore and deconstruct gender performativity in a digital space. While this study highlights the social construction of gender identity among Replika users, it is limited by the scope of its qualitative analysis and the absence of quantitative data. Future research could expand on this foundation by employing mixed-method approaches, incorporating statistical analysis, and exploring additional contexts to deepen the understanding of how AI technologies interact with and shape perceptions of gender identity.

References

- Ahn, J., Kim, J., & Sung, Y. (2022). The effect of gender stereotypes on artificial intelligence recommendations. *Journal of Business Research*, 141, 50-59.
- Bernotat, J. (2021). Keep an Eye on Stereotypes—The Impact of Gender Stereotypes (Toward Humans and Robots) on Language Processing (Doctoral dissertation, Dissertation, Bielefeld, Universität Bielefeld).
- Bolt, B. (2023). Performance, Performativity, and Subjectivity at the Intersection of Art and Digital Cultures. In *Choreomata*(pp. 95-121). Chapman and Hall/CRC.
- Butler, J. (1993). Poststructuralism and postmarxism. *diacritics*, 23(4), 3.
- Butler, J. (1999). Revisiting bodies and pleasures. *Theory, culture & society*, 16(2), 11-20.

- Costa, P., & Ribas, L. (2019). AI becomes her: Discussing gender and artificial intelligence. *Technoetic Arts: A Journal of Speculative Research*, 17(1-2), 171-193.
- Craiut, M. V., & Iancu, I. R. (2022). Is technology gender neutral? A systematic literature review on gender stereotypes attached to artificial intelligence. *Human Technology*, 18(3), 297-315.
- Hine, C. (2020). *Ethnography for the internet: Embedded, embodied and everyday*. Routledge.
- Laestadius, L., Bishop, A., Gonzalez, M., Illenčik, D., & Campos-Castillo, C. (2024). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society*, 26(10), 5923-5941.
- Leavy, S. (2018). Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. In *Proceedings of the 1st international workshop on gender equality in software engineering* (pp. 14-16).
- Nass, C., Moon, Y., & Green, N. (1997). Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers with Voices. *Journal of Applied Social Psychology*, 27(10), 864 - 876
- Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140, 107600.
- Possati, L. M. (2023). Psychoanalyzing artificial intelligence: The case of Replika. *AI & SOCIETY*, 38(4), 1725-1738.
- Ringrose, J., & Rawlings, V. (2015). Posthuman performativity, gender and “school bullying”: Exploring the material-discursive intra-actions of skirts, hair, sluts, and poofs. *Confero: Essays on education, philosophy and politics*, 3(2), 80-119. <https://doi.org/10.3384/confero.2001-4562.150626>
- Thellman, S., Hagman, W., Jonsson, E., Nilsson, L., Samuelsson, E., & Simonsson, C. et al. (2018). He is not more persuasive than her. *Proceedings of the 18Th International Conference On Intelligent Virtual Agents*.
- Weber-Guskar, E. (2022). Reflecting (on) Replika. *Social Robotics and the Good Life*, 103.
- Wellner, G. P. (2020). When AI Is Gender-biased. *HUMANA. MENTE Journal of Philosophical Studies*, 13(37), 127-150.
- Xie, T., & Pentina, I. (2022). Attachment theory as a framework to understand relationships with social chatbots: a case study of Replika.
- Xie, T., Pentina, I., & Hancock, T. (2023). Friend, mentor, lover: does chatbot engagement lead to psychological dependence?. *Journal of service Management*, 34(4), 806-828.
- Yoon, H. (2021). Digital flesh: A feminist approach to the body in cyberspace. *Gender and Education*, 33(5), 578-593.