User Profiling and Market Segmentation of Virtual Idols: A Cluster Analysis Based on Virtual Idol Liu Yexi

Qihan Lin*

Xiangsihu College of Guangxi Minzu University, Nanning 530225, China

*Corresponding email: 2504575390@qq.com

Abstract

With the rapid advancement of digital technology and social media, virtual digital humans and hyper-realistic virtual idols have demonstrated immense commercial potential and application value. Taking virtual idol Liu Yexi as a case study sample, this research employs a multi-faceted approach integrating literature review and questionnaire surveys to delve into audience characteristics and interaction preferences. Employing a two-stage clustering algorithm combining Self-Organizing Map (SOM) neural networks with K-means clustering algorithm, the audience is segmented into four distinct groups: the highly educated and rational type prioritizing logic and aesthetics; the young and fervent type exhibiting strong emotional dependency; the mature and pragmatic type with demanding requirements; and the socially driven type seeking recognition. Based on these findings, user personas are constructed centered around four primary markets: the core emotional monetization market, the premium content experience market, the long-term Intellectual Property (IP) oversight market, and the broad social viral market. These markets are further segmented. The research aims to reveal the group characteristics of different market segments, providing guidance for developing refined operational and targeted marketing strategies for virtual digital human IPs. This seeks to advance audience expansion and commercialization within the virtual digital human industry.

Keywords

Virtual idols, User profiling, Market segmentation, Cluster analysis, Liu Yexi

Introduction

With the rapid advancement of digital technology and social media, virtual digital humans have emerged as brand representatives, achieving broader application across marketing scenarios and generating commercial value. Currently, virtual digital humans can be categorized into two primary types: human-driven and Artificial Intelligence (AI)-driven. Human-driven models primarily utilize motion capture technology to animate virtual digital humans, while AI-driven models rely on hyper-realistic three-dimensional (3D) technology for presentation [1]. Current virtual digital humans are evolving towards hyper-realistic virtual idols, satisfying users' emotional needs through anthropomorphic traits. Furthermore, with the rapid development of the metaverse, virtual digital humans have garnered attention from major brands as a bridge connecting the metaverse and the real world. Notable brands such as Huaxizi and Mercedes-Benz have successively launched their own brand virtual humans.

On October 31st, 2021, "Liu Yexi" released her inaugural video via her Douyin account, showcasing virtual makeup application that drew considerable user engagement. Employing suspenseful storytelling, beauty techniques, narrative elements, and post-production effects, her debut video garnered 3.00 million likes and attracted over a million followers. Adhering to a philosophy of original content and technology, Liu Yexi crafts technologically sophisticated short videos, rendering each production akin to a cinematic blockbuster.

Liu Yexi's success stems both from sophisticated technical execution and diverse audience support, making precise exploration of her audience's characteristics and differing needs crucial. Against this backdrop, this paper selects virtual idol Liu Yexi as its research subject. By mining user interactions and

questionnaire data from relevant social platforms and applying cluster analysis, it constructs user personas for her audience before conducting market segmentation. This paper aims to reveal the characteristics and preferences of user groups across different segments, thereby providing targeted, refined operational and marketing strategies for virtual digital humans. This holds significant importance for expanding the audience reach and facilitating the commercialization of the virtual digital human industry.

Literature review

Domestic literature primarily addresses virtual idol marketing, consumer behavior, and brand image management. It offers extensive theoretical analysis and empirical research that underscores the importance of virtual idols in cultural dissemination and brand marketing. The intrinsic mechanisms and advancement strategies of virtual idol marketing are examined from a macro perspective, and their contemporary value is also explored, revealing the potential and challenges of virtual idols in brand marketing. The impact of virtual idol marketing on consumer behavior is another focus, showing how virtual idols shape consumer attitudes and purchase intentions. Technology-enabled quasi-social interactions of virtual idols are explored, analyzing the relationship between virtual idols and their fanbase. Pathways for virtual digital humans to empower the popular culture industry are analyzed from the perspectives of convergence, innovation, and transformation. Young people's worship of virtual idols and the associated moral dilemmas in the digital age are investigated, with guiding strategies proposed. Pathways for virtual idols to disseminate traditional music culture are explored, emphasizing their role in cultural inheritance. The acceptance and influence of virtual streamers among Generation Z are investigated, providing theoretical support for understanding virtual idols' societal impact. The foundations and formation process of virtual idols as emerging idol imaginaries are analyzed, exploring their sociocultural significance. The influence of virtual idol endorsers on Generation Z consumers' purchase intentions is investigated, revealing the mediating role of consumer brand identification [2].

International literature places increasingly greater emphasis on virtual idols' social interactions, fan cultures, and psychological motivations, thus offering quite fresh perspectives on their global influence. Fan et al. conducted a comprehensive experimental study comparing young consumers' responses to advertisements featuring virtual idols, online influencers, and traditional celebrity endorsements, providing empirical data on the effectiveness of different celebrity endorsement types [3]. Chen and Skey employed Collins's interactive ritual chain model to examine the culture of Chinese virtual idol fandom, exploring how fans establish connections through content production [4]. Onishi et al. investigated interactions between robotic social tactile interfaces and virtual idols, analyzing the embodied nature, visibility, and politeness of virtual idols [5]. Song et al. examined mainstream culture's influence on virtual idol online fan communities, analyzing the relationship between virtual idols and youth identity [6]. Wang et al. analyzed virtual idol fans' creative behaviors from a psychological perspective based on Motivation, Opportunity, Ability (MOA) theory [7].

Although academia has extensively explored the virtual digital human and virtual idol industries, existing research predominantly focuses on macro-level development trajectories and the evolution of business models. It rarely engages in micro-level empirical investigations into the internal diversity of virtual idol audiences, user profiling, or market segmentation. Concurrently, research methodologies predominantly rely on theoretical deduction and qualitative case descriptions, lacking targeted exploration of audience characteristics through quantitative research methods such as cluster analysis based on real user data.

Defining the concept of virtual idols

The concept of virtual idols can be traced back to the 1980s, when technology was still relatively primitive, relying primarily on simple computer graphics and animation techniques. In 1982, Japan introduced the representative virtual idol Mingmei Hayashi in the TV anime Macross. However, due to technological limitations, her influence remained limited. During the 1990s, advancements in computer graphics technology paved the way for the emergence of virtual idols. The concept and technology began to develop and gain

traction in the United States around 1990. Progress in computer graphics, animation techniques, and voice synthesis provided the technical foundation for virtual idols. By the early 21st century, widespread internet adoption and rapid advancements in computer graphics technology ushered in a period of significant technical breakthroughs for virtual idols. In 2007, Japan introduced the virtual idol "Hatsune Miku", whose voice was generated using Vocaloid software and whose appearance was created through 3D modeling technology. She quickly gained immense global popularity. Hatsune Miku's success stemmed from two key factors: technological innovation and an open creative model. This model allowed fans and creators to produce songs and animations for her, significantly expanding her influence. By 2010, virtual idols entered a phase of diversified development. Continuous technological advancements, including motion capture, real-time rendering, and artificial intelligence, made their performances more lifelike and natural. Simultaneously, the types of virtual idols and their application scenarios grew increasingly diverse, appearing everywhere from music and dance to live streaming and film. Furthermore, the commercial value of virtual idols was fully tapped, with diverse business models emerging, including brand endorsements, concerts, and merchandise.

Thus, virtual idols can be defined as an innovative cultural phenomenon. They fuse advanced virtual digital technology with deep-rooted otaku culture and idol worship concepts, manifesting across mediums like painting, animation, and computer-generated imagery. They possess the ability to interact across the boundary between virtual and real worlds. Virtual digital humans, merging the essential characteristics of virtual and digital humans, manifest as 3D simulated figures with novel personality traits. They maintain a certain similarity to real humans while potentially serving as digital mirror images of actual individuals, digital twin humans. This technology enables the instant creation of virtual idol IPs and facilitates boundless expansion [8].

Data collection and cluster analysis

Data collection and pre-processing

The questionnaire primarily targeted audiences associated with virtual idols. Participants completed the survey via a link on the Wenjuan Xing website (https://www.wjx.cn). Respondents were required to have prior interaction experience with virtual idols, with the questionnaire distinguishing between those who had and those who had not. Multiple rounds of sample screening were conducted to ensure high relevance of participants' age, educational attainment, and occupation to user characteristics. Questionnaire One distributed 414 questionnaires in total, with 401 valid responses collected.

The integrity and scientific validity of the questionnaire data underwent comprehensive verification, including checks for missing values. No missing data were identified across variables such as age, occupation, gender, and income. Key data points including income, age, and interaction methods were standardized. A formula was applied to ensure the relative weighting of each variable in the cluster analysis remained reasonably balanced, thereby establishing a comprehensive cluster analysis indicator model. The standardization formula is as follows.

$$Z = \frac{X - \bar{X}}{s} \quad (1)$$

Cluster Analysis

Given the traditional K-means algorithm's sensitivity to initial cluster centers and outliers, this study employs a two-stage clustering method combining self-organizing mapping neural networks with K-means. The aim is to more accurately characterize the user profile of virtual idol Liu Yexi and conduct scientific market segmentation. Compared to single K-means clustering, this approach utilizes the SOM network to pre-map high-dimensional user feature data onto a low-dimensional two-dimensional grid. This not only effectively reduces data dimensions and filters out noise but also preserves the local topological features of the original user data well, thereby providing a more robust initial feature space for subsequent market segmentation.

In the implementation process, the acquired Liu Yexi user feature data first undergo preprocessing. Missing values are imputed using the feature mean method, and dimensional effects are eliminated through standardization based on standard deviation. Subsequently, a 10.00×10.00 two-dimensional SOM network topology is established for preliminary training,

with an initial learning rate set at 0.50. During computation, the distance between the input user sample vector x and the weight vector w_j of each neuron in the competitive layer is calculated to identify the winning neuron with the smallest distance. The Euclidean distance formula is:

$$d_j = \sqrt{\sum_{i=1}^{n}(x_i - w_{ji})^2} \quad (2)$$

Where x_i denotes the i th dimensional feature of the input data point x ,w_{ji} represents the corresponding weight for the j th neuron, and n indicates the feature dimension of the data point. Upon determining the winning neuron, the network updates the weights of this neuron and its topologically neighboring nodes over 1000 random iterations using the following formula.

$$w_{ji}(t+1) = w_{ji}(t) + \eta(t) \cdot h_{c,j}(t) \cdot (x_i - w_{ji}(t)) \quad (3)$$

In the formula, t denotes the iteration count, $\eta(t)$ represents the monotonically decreasing learning rate, and $h_{c,j}(t)$ is the neighborhood function for the winning neuronc .

Following comprehensive mapping by the SOM network, the two-dimensional grid coordinates of the SOM winning neuron corresponding to each user data point are extracted as input for secondary clustering. Subsequently, the K-means algorithm is applied to cluster these reduced-dimensional two-dimensional coordinates. Various K values (ranging from 2 to 9) are tested. The contour coefficient is calculated for each K value as an evaluation metric. Analysis reveals that when K=4, the contour coefficient achieves its maximum value, demonstrating excellent performance with significant inter-cluster differentiation and compact intra-cluster organization. Ultimately, while are partitioned setting a random seed to ensure reproducibility, the audience segments for Liu Yexi were consistently partitioned into Groups A, B, C, and D based on the K-means clustering of the SOM mapping coordinates. This two-stage clustering strategy not only enhances the algorithm's robustness when processing complex user behavior data but also endows each segmented user group with stronger practical interpretability.

Analysis of clustering results

(1) Overview of clustering results

Based on the K-means clustering of SOM mapping coordinates, 401 valid samples were categorized into four distinct groups with pronounced characteristics. Group A (corresponding to Cluster 0, 99 individuals) prioritizes content logic and aesthetics, exhibiting high educational attainment and income levels, with exceptionally rational consumption decisions. Group B (corresponding to Cluster 1, 101 individuals) exhibits high activity levels, strong emotional attachment, and a youthful demographic. They demonstrate the strongest willingness to spend out of devotion, forming the core driving force of interactions. Group C (corresponding to Cluster 2, 102 individuals) comprises mature, pragmatic consumers who demand rigorous long-term IP management and consistent character portrayals. Their purchasing preferences are grounded in brand reputation and product quality. Group D (corresponding to Cluster 3, 99 individuals) consists of socially driven audiences with relatively low interaction frequency. They exhibit high dependence on fan community belonging and are significantly influenced by their social circles in consumption decisions.

(2) User profile dimensions

From a demographic perspective, gender differences may influence audience preferences for virtual idols and consumption behaviors. Age impacts audience interests and spending power, while occupation and educational attainment correlate with economic income, interaction backgrounds, and engagement methods. Collectively, these factors shape audience choices regarding virtual idol interactions and their experiential demands.

Behavioral analysis reveals that interaction frequency indicates audience affection and sustained engagement. Key factors prioritized during interactions, such as character design, persona depth, and content authenticity, reflect core experiential pursuits. Participation in co-creation activities stems from emotional projection and self-actualization needs, while purchasing endorsed products directly manifests commercial conversion logic.

(3) User profiling for different audience segments

a) Group A: Highly educated, rational content fans prioritizing logic and aesthetics

This group exhibits the highest average educational attainment (3.42) and relatively high monthly income (3.07) among all samples, representing a quintessential highly educated, rational demographic. Regarding pain points, they exhibit exceptionally stringent demands for

virtual idol content quality and logical consistency. A substantial 76.77% perceive the primary issue as mismatched character designs, while 73.74% highlight a sense of inauthenticity in creative output. In terms of co-creation participation and consumption motivation, they demonstrate strong expressive desires, showing highest willingness to engage in Liu Yexi's textual design (62.63%). When converting to consumption, they demonstrate remarkable rationality, avoiding impulsive purchases and prioritizing the brand image and reputation of endorsed products (60.61%) alongside the product's inherent quality (50.51%).

b) Group B: Young, passionate co-creators with strong emotional dependency

This group constitutes the core audience of Liu Yexi's IP, predominantly comprising young individuals. It is the youngest (age 3.25 years old) and has the lowest average monthly income (2.46) among the four clusters. They exhibit an exceptionally high demand for interaction with Liu Yexi, being most intolerant of limited interaction formats (76.24%) and insufficient interaction depth (67.33%). They are also the only group citing emotional connection with Liu Yexi (70.30%) as their primary motivation for engagement. Their co-creation enthusiasm is exceptionally high, with strong willingness to participate deeply in textual design (78.22%) and character development (53.47%). Concurrently, they demonstrate the most proactive consumption intent, being most likely to purchase products directly due to affection and trust in Liu Yexi (50.50%), forming the core foundation for commercial monetization.

c) Group C: Discerning, mature, pragmatic career-focused fans

This group features a higher average age (3.53) and income level (3.03), possessing considerable cultural literacy and life experience. They impose stringent demands on the operational pace of virtual idols, most critically citing slow avatar update frequency (65.69%). They also keenly identify deeper operational issues with Liu Yexi, including insufficient exploration of character depth (66.67%) and lack of character stability (66.67%). This group is most enthusiastic about reshaping Liu Yexi's character design (69.61%). Notably, this group exhibits the lowest proportion of purchases driven by affection and trust (23.53%), demonstrating entirely rational consumption decisions heavily reliant on brand image and reputation (60.78%).

d) Group D: Socially driven community followers seeking recognition

This segment exhibits relatively low engagement, with the highest proportion of female members (1.45), the oldest age group (3.65 years old), and lower income levels. They prioritize the acceptance of fan opinions, with 63.64% complaining about the lack of official consideration for fan feedback. They also perceive insufficient application of current technology, particularly in the presentation of skins (77.78%). Their behavioral logic is highly dependent on social attributes, with the core reason for participation being the search for a sense of belonging within the fan community (69.70%). In shopping decisions, they are highly susceptible to social influence (50.51%).

Market segmentation and demand analysis

Based on the above profile, Liu Yexi's audience market can be segmented as follows.

(1) Core emotional monetization market

Within this market, young individuals form the dominant demographic, characterized by vitality and an exploratory spirit. They seek projects incorporating modern elements and distinctive creativity. This audience exhibits both high creative engagement and emotionally driven purchasing power, with the majority willing to pay for what they cherish. This represents a high-value emotional market most receptive to virtual idol IP merchandise, blind boxes, and co-branded consumer goods [9].

(2) Premium content experience market

This market's audience demands exceptional visual alignment and content authenticity. The premium consumption segment holds significant potential, with customized services and high-caliber events enhancing premium pricing opportunities. They exhibit extreme sensitivity to IP reputation, requiring sustained output of cinematic-quality short videos and meticulously crafted narrative logic to maintain engagement.

(3) Long-term IP supervision market

Within this segment, audiences seek to experience an IP's depth and consistency through interaction. They exhibit low tolerance for shallow character development

and sluggish updates. The market's consumption potential lies in cross-industry collaborations with premium physical brands, focusing on the product's value proposition and brand ethos [10].

(4) Social virality market

This market's audience prioritizes foundational interactive services and favors acquiring social currency through fan communities. Positioned as a broad social entertainment market, fulfilling this audience's sense of belonging can significantly boost Liu Yexi's secondary dissemination rates and topic popularity across major social platforms.

Operational recommendations for virtual idol Liu Yexi

Precision marketing to align with market demands

IP operators should tailor differentiated marketing strategies based on segmented market characteristics. For young, passionate co-creation fans (Group B), emphasize emotional companionship by launching virtual merchandise with blind box elements or exclusive digital collectibles for fans. For mature, pragmatic career-focused fans (Group C), commercialization should involve selected brand endorsements. Moreover, cross-industry partnerships with high-quality, high-profile physical brands should be pursued, leveraging product quality to stimulate purchasing power. For socially-driven community fans (Group D), harness the traffic advantages of integrated media platforms to launch viral activities such as Liu Yexi lookalike makeup challenges and fan-created content competitions, fulfilling their social and belonging needs.

Technological innovation to meet diverse demands

In the realm of technological application and innovation, the operations team must closely address the specific pain points of different market segments. Group A has stringent demands for skin-suit accuracy, and Group D has expectations for technical presentation. To address these, continuous adoption of advanced motion capture and rendering technologies is essential to refine micro-expressions and enhance the realism of the skin-suit's physics engine. For Group B's intense desire for deep real-time interaction, exploring AI technologies like large language models should be pursued. This would grant Liu Yexi more immediate and personalized conversational capabilities, overcoming the current technical limitation of predominantly one-way short-video output.

Refined operations to enhance fan retention

Liu Yexi must deepen long-term IP character development. Addressing Group C's concerns about insufficient character depth and instability, the writing team should construct a more expansive and internally consistent metaverse worldview, solidifying Liu Yexi's core fusion of traditional Chinese aesthetics and cyberpunk elements. Concurrently, a robust fan feedback oversight mechanism should be established. Addressing Group D's concern over insufficient fan input adoption, official accounts could regularly host face-to-face planning sessions or plot-voting activities. This transforms fans from passive observers into collaborative partners in IP development, thereby comprehensively sustaining and enhancing Liu Yexi's long-term viability and commercial value [11].

Conclusion

This study conducts a cluster analysis of virtual idol Liu Yexi's audience, segmenting users into four distinct groups: highly educated rational content fans (Group A), young passionate co-creators with strong emotional dependency (Group B), mature pragmatic career-focused fans (Group C), and socially driven community followers (Group D). Based on these profiles, four corresponding markets are identified: the core emotional monetization market, the premium content experience market, the long-term IP supervision market, and the broad social viral market. The findings reveal that different audience segments exhibit heterogeneous consumption logics, interaction preferences, and pain points, necessitating differentiated operational strategies. To sustain long-term IP value, operators should implement precision marketing aligned with each segment's demands, continuously advance technological innovation to enhance realism and interactivity, and refine fan engagement mechanisms to foster co-creation. These insights provide actionable guidance for virtual idol IP management and contribute to the commercialization and audience expansion of the virtual digital human industry.

Funding

This work was not supported by any funds.

Acknowledgements

The author would like to show sincere thanks to those techniques who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Phua, J., Evans, N. J., Ko, Y., Lee, J. (2025) Can virtual, CGI-generated, influencers help sell products on Instagram? Effects of perceived realism and disclosure on brand-related attitudes and behavioral intentions. *International Journal of Internet Marketing and Advertising*, 23(3), 233-254.

[2] Yu, Y., Bannasilp, A., Kwong, S. C. M. (2025) Virtual idols' influence on consumers' brand attitude and purchase intention: a perspective of para-social interaction. *Electronic Commerce Research and Applications*, 74, 101546.

[3] Fan, F., Fu, L., Jiang, Q. (2024) Virtual idols vs online influencers vs traditional celebrities: How young consumers respond to their endorsement advertising. *Young Consumers: Insight and Ideas for Responsible Marketers*, 25(3), 329-348.

[4] Chen, Z., Skey, M. (2024) "I produce songs for her …. In this way, I gradually know her more. The more I know her, the more I like her": Using Collins' model of interactive ritual chains to study the case of virtual idol fandom in China. *Convergence: The International Journal of Research into New Media Technologies*, 30(2), 841-859.

[5] Onishi, Y., Ogawa, K., Tanaka, K., & Nakanishi, H. (2024) Embodied, visible, and courteous: Exploring robotic social touch with virtual idols. *Frontiers in Robotics and AI*, 11, 1240408.

[6] Song, L., Chow, O. W., Na, M., Sarjit, S. G. (2024) Virtual idol and youth identity: the impact of mainstream culture on the virtual idol online fanbase community in China. *Studies in Media and Communication*, 12(2), 100-115.

[7] Wang, Q., Long, S., Zeng, Y., Tang, L., Wang, Y. (2023) The creative behavior of virtual idol fans: a psychological perspective based on MOA theory. *Frontiers in Psychology*, 14, 1290790.

[8] Ju, N., Kim, T., Im, H. (2024) Fake human but real influencer: The interplay of authenticity and humanlikeness in virtual influencer communication? *Fash Text*, 11(16), 1-27.

[9] Thakur, A., Singla, K., Singla, K., T., M. I., Kaur, S., Kaur, P. (2025) Beyond the pixels: psychographic profiling of virtual influencersâ€™ followers in social media. *Indian Journal of Marketing*, 55(3), 52-68.

[10] Xiang, D., Yin, Y., Ge, M., Wang, Z. (2025) The impact of brand-developing versus collaborative virtual influencer endorsement selection strategies on consumer engagement. *Advances in Psychological Science*, 33(6), 965-983.

[11] Kim, E. (Anna), Shoenberger, H., Kim, D., Thorson, E., Zihang, E. (2025) Novelty vs. trust in virtual influencers: exploring the effectiveness of human-like virtual influencers and anime-like virtual influencers. *International Journal of Advertising*, 44(3), 453-483.