

Visual Representations in AI Mini Dramas

AI-Assisted Annotation and Analysis of AI-Generated Visual Narratives

Team Members

Facilitators:

Dan danxu0910@gmail.com

Piera p.riccio@uva.nl

Carlo carlo.dgtn@gmail.com

Sabine s.m.c.niederer@hva.nl

Inte i.gloerich@hva.nl

Participants:

Rebekah Lyndon rglyndon@protonmail.com

Giulia Ratti ratti.giulia00@gmail.com

Morgan Park mpark@cuhk.edu.cn

Lulu YUAN luluyuan@life.hkbu.edu.hk

Jing Cai cjjingcai@gmail.com

Yilin Tang yt413@cam.ac.uk

Andréa Nicolas andrea.nicolas@student.uva.nl

Anna Ivanova a1vanova@tlu.ee

Rinlapas Ketverapong r.ketverapong.1@research.gla.ac.uk

Design facilitator:

Anna Cattaneo anna.cattaneo@polimi.it

Links

Poster:

https://drive.google.com/file/d/15S8IXf0q23RBSbNeVujzEn7yWMvwglGx/view?usp=drive_link

AI Baby Drama Data Set:

https://drive.google.com/drive/folders/1scd9ly9o9eeRh5A6NQTIU2Rq4wgDsQA6?usp=drive_link

Contents

Team Members	1
Contents	1
Summary of Key Findings	2
1. Introduction	2
2. Initial Data Sets	2
3. Research Questions	2
4. Methodology	2
5. Findings	2
6. Discussion	2
7. Conclusion	2
8. References	2

Summary of Key Findings

- AI constructs rather than merely reflects gender representation. Although fruit avatars are inherently gender-neutral, AI-generated mini-dramas consistently transform them into gendered symbols through visual design, narrative roles, and dialogue. This process reinforces conventional associations of masculinity with strength and agency, and femininity with attractiveness, emotional vulnerability, and relationship-oriented roles.
- The AI baby is not merely a cute digital novelty but a culturally loaded figure through which Douyin mediates and transforms the social life of its users. Cuteness, in this context, functions as an affective technology: it renders labour precarity comic, family obligation sentimental, ethnic difference adorable, and human-machine boundaries playful.

1. Introduction

AI mini-dramas are a rapidly growing form of short-form, AI-generated visual storytelling distributed through platforms such as TikTok, Douyin, YouTube Shorts, and Instagram Reels. They form part of the broader ecosystem of AI slop, characterized by low-cost, high-volume, attention-driven content. Although presented as light entertainment, these narratives frequently rely on exaggerated social, cultural, and gender stereotypes.

This project applies an AI-assisted workflow for annotating and analyzing AI mini-dramas. Using vision-language models, we automatically identify and describe characters, visual cues, actions, and social roles across a selection of AI-generated videos. Specifically, we looked into two case studies – AI fruit drama and AI baby drama – as first examples to test and evaluate this workflow.

2. Initial Data Sets

Case study 1: AI Fruit drama

Searching within Instagram leads to more algorithmically personalised results. To somewhat mitigate this, we initially remained logged out of Instagram to search for videos related to “fruit drama” that Instagram categorises as popular, following this url: <https://www.instagram.com/popular/fruit-drama>. While logged out, the videos displayed appeared consistent across browsers on 31st June 2026, accessed from the Netherlands.

Though all those listed had view numbers at least in the hundreds of thousands, to ensure focus on videos with high visibility we then manually selected the videos with at least one million views. We then collated these by opening each in a new tab, then logging into Instagram, running Zeeschuimer and navigating to each tab in sequence. In the end, 16 videos were collected and analyzed.

Case study 2: AI Baby drama

AI baby videos have recently become a highly viral trend on Douyin in China. These videos feature hyper-realistic AI-generated or AI-filtered babies who are placed in everyday scenarios, such as singing, doing housework, participating in casual conversations, or imitating adult speech and behaviour. Through these scenes, AI babies are not merely cute digital figures. They generate new forms of performativity through familiar social roles and emotional interactions, while reconfiguring cuteness in an uncanny way that blurs the boundaries between infancy, adulthood, and artificial generation.

This study then approaches AI baby videos as a platform-specific form of synthetic affective media. The AI baby is not only represented as an innocent or adorable child, but is repeatedly made to work, study, care, perform, speak, and embody cultural identities. In this sense, the baby body becomes a flexible digital figure through which adult labour pressure, family intimacy, ethnic imagination, and platform entertainment are translated into cute and shareable short-video forms. The apparent lightness of these videos therefore carries a more complex social and affective function: cuteness becomes a way of making pressure, obligation, difference, and artificiality emotionally accessible.

Searching within Douyin can produce results shaped by platform recommendation and user-personalized ranking. To make the sampling process more transparent, we searched Douyin using the Chinese keyword “**AI小孩**” (*AI child*), a commonly used term for AI-generated baby and child videos on the platform. The search was conducted through keyword- and hashtag-based browsing, focusing on videos that Douyin presented as highly visible within this topic area.

Platform objects were captured through Zeeschuimer during the browsing process and then imported into 4CAT for dataset organization, metadata export, and initial filtering. The exported metadata included video ID, post URL, author, caption, hashtags, upload information, and engagement indicators. To focus on videos with strong platform visibility and circulation, we selected the [top 15 videos with the highest view counts](#)

from the collected dataset as the close-analysis sample. These 15 AI reborn baby videos were then downloaded and analyzed through multimodal coding.

3. Research Questions

RQ1: How are social roles, identities, and stereotypes visually constructed and represented in AI mini dramas?

RQ2: What are the affordances and limitations of AI-assisted annotation for studying representation in AI-generated visual media?

RQ for Fruit Drama:

1. How are gender, identities and social role are visually constructed and represented in Ai-generated fruit mini-drama
2. How narrative roles reinforce gender representation and stereotypes in Ai-generated fruit mini-drama
3. How dialogue contributes to the construction of binary gender representation in AI fruit mini-drama.

RQ for Baby Drama:

1. What recurring patterns of representation can be identified in AI baby videos on Douyin?
2. How are AI baby bodies made to perform social roles, everyday pressures, and emotional relations, and what affective work does this infantilization perform?
3. How is cuteness produced, circulated, and transformed through AI-generated infant bodies?

4. Methodology

Video collection

16 Instagram videos collected via "fruit drama"/"fruitdrama" queries (1M+ views);
15 Douyin videos collected as top results for "ai baby"

Data prepping

Frames extracted at each scene change via Python, audio transcribed with Whisper, transcript split and aligned to detected scenes

Human reflections

AI-facilitated aggregation of clusters of interest (e.g. characters, children) was reviewed and combined with human analysis based on viewing the videos

Codebook design

Manually drafted codebooks, theme-specific, refined by AI for structure and prompt use, then human-validated

Assisted annotation

Codebook and prepped data fed to AI models (GPT, Claude) to perform coding based on the defined variables

Fig 1. An illustration of the workflow

Both case studies follow the same AI-assisted workflow after data collection. Following the procedures described in the dataset section, we first developed theme-specific codebooks, which were refined with the assistance of AI tools to improve their structure and generate annotation prompts. The resulting codebooks were then manually reviewed and validated.

To prepare the datasets, we extracted keyframes from each video using Python-based scene detection and transcribed the accompanying speech using OpenAI's Whisper model. The keyframes, transcripts, and codebook prompts were subsequently processed by general-purpose AI models to annotate based on predefined categories and identify recurring patterns and clusters across the data.

Finally, all AI-generated annotations and clusters were manually reviewed against the original videos to verify their accuracy, refine interpretations, and assess the strengths and limitations of the AI-assisted analysis.

- Codebook design:
 - Fruit drama: A structured codebook was developed to systematically annotate AI-generated fruit dramas **at the character and scene level**. For each scene, the codebook captures character identity and continuity, fruit type, gender presentation and its visual and audio cues, age, social relationships, narrative archetypes, profession, actions, setting, time period, and notable objects. The annotation categories were designed to combine **descriptive visual features with narrative and social dimensions**, enabling both quantitative comparison across videos and qualitative analysis of recurring representational patterns.
 - Baby drama: For the AI Baby Drama case study, the codebook was designed around three layers: **descriptive visual coding, analytical code families**, and **tech-political memo writing**. The descriptive layer records basic metadata and visual features for each video, including top-view rank, video ID, post URL, author, caption, engagement data, action, background/scenario, character or social role, dress/costume, body appearance, props, subtitles, and visual cues. The second layer organizes these observations into four code families: **narrative codes, aesthetic codes, scenario codes**, and **body/affect codes**. These capture recurring patterns such as Gen Z labor irony, intergenerational care, ethnicized comparison, screen culture, platform performance, hyper-cute infant aesthetics, uncanny/reborn-doll realism, domestic settings, hospital scenes, multicultural montages, adult-role performance, dancing bodies, and cute bodies carrying adult topics. The final layer consists of

tech-political analytic memos and coder-confidence notes, which connect descriptive codes to broader arguments about AI-generated infant bodies, cuteness, social pressure, domestic intimacy, ethnicized templates, and platformized performance.

- Data preparation:
 - Fruit drama: Video files are first segmented into scenes using the **PySceneDetect** library with the ContentDetector algorithm, which identifies shot transitions and extracts a representative keyframe for each scene. In parallel, audio is transcribed locally using OpenAI's **Whisper** model, which also performs automatic language detection. The resulting transcripts are structured into a scene-aligned CSV file containing start and end timestamps and corresponding text, and are linked to the extracted frames through a shared identifier combining the video ID and scene number.
 - Baby drama: The selected videos were then downloaded, organized by video ID, transcribed through **Whisper/faster-whisper**, and processed through **FFmpeg/Python scripts** for key-frame extraction. Using **byjlw/video-analyzer** as a reference pipeline, metadata, transcripts, and extracted frames were structured into analyzable outputs.
- Automated annotation:
 - Fruit drama: Frames and scene-aligned transcripts were processed jointly by **Claude** in batches of videos. The model analyzed each video as a continuous narrative, identifying characters across scenes with consistent IDs and linking visual information with transcribed dialogue. When needed, it cross-checked captions and transcripts for alignment and relied on visual cues when audio was unavailable or unclear, enabling coherent annotation of characters and narrative roles across entire videos.
 - Baby drama: The final preliminary coding was conducted with **Codex/GPT-5**, which generated visual and multimodal codes from the extracted frames, transcripts, and metadata. These included codes for action, dress, background, social role, body appearance, props, facial expression, visual cues, narrative, aesthetics, scenario, and body/affect.
- Human evaluation:
 - Fruit drama: Each of the fruit videos was viewed by participants, to allow for closer and more holistic consideration of the content. Through discussion, we selected characters that best illustrated key tropes and common themes in the narratives. Our visual analysis of the aesthetic presentation of these characters and understanding of their roles in the

stories allowed us to assess the messaging of the videos with greater nuance. Using the AI-facilitated grouping of individual characters, we then reviewed the still frames of the videos to select an image per figure, focussing on instances where characters displayed emotive expressions. We reviewed and updated the categorisation of the characters by AI to reflect our understanding of the characters, grouping them as victims, abusers, allies, and mixed roles – allowing us to preserve some of the ambiguity present in some characterisations. The AI's gendering of characters largely aligned with our human understanding of them. However, the AI output for “gender_visual_cues” remained distinct from our reasoning for character genders. Cues included objects that in themselves are not gendered (“sunglasses”, “police uniform”, “hospital gown”), raising questions about the processes behind the categorisations and their justifications. Visual cues also referenced behaviours such as “tearful” – would be interesting to know more about the basis for this. etc

- Baby drama: For the AI Baby Drama case study, all AI-generated codes were reviewed, corrected, and reorganized by the researcher. The researcher checked the accuracy of visual descriptions, refined the analytical codes, and added tech-political memos for each video. The reviewed codes were then organized into coding sheets, frequency tables, co-occurrence tables, dashboards, and heatmaps. Each entry also included a coder-confidence level to indicate the stability of the coding decision after researcher review. For the final analysis, the codes were grouped into four interpretive categories: **Aspirational Laboring Baby**, **Familial Care Baby**, **Ethnicized Slop Baby**, and **Platformized Performing Baby**. These categories show how AI baby videos use infant bodies to make adult pressure cute, re-stage domestic intimacy, standardize cross-cultural ethnic templates, and generate uncanny cuteness through AI-mediated platform performance.

5. Findings

Describe your findings. Consider any counter-intuitive findings.

AI fruit drama

betrayed partner

mis-treated employee

un-favored sibling

 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	 neglectful mother	 aggressive mother	 exploitative boss	 deceiving sibling	 reckless criminal	 violent husband	 manipulative partner	
--	---	--	--	--	---	--	--	---	--	--	--	--	--	--	---	--	--	--	--	--	--	---	--	--	--	--	--	--	---	--	--	--	--	--	--	---	--	--	--	--	--	--	---	--	--	--	--	--	--	---	--	--	--	--	--	--	---	--	--	--	--	--	---	--	---	---	---	---	---	---	--	--

recruiting doctor	
stronger	
fair employer	
exposer	
comforting mother	
kid hero	

 A row of 12 images of various food-based characters from the movie 'The VeggieTales'. From left to right: a small yellow chick (a favored sibling), a pink pig (a piglet), a brown bear (a bear), a red chili pepper (a chili), a green pea (a pea), a brown mushroom (a mushroom), a red tomato (a tomato), a green cucumber (a cucumber), a green pea (a pea), a green pea (a pea), a green pea (a pea), and a green pea (a pea).

Gender is visually constructed through fruit avatars

Appearance emerged as the primary mechanism through which gender was communicated. Even when characters changed gender across different scenes, gender identity was represented through external visual markers such as wigs, clothing, or make-up rather than through ambiguous or fluid embodiments. Consequently, gender was consistently presented as binary and highly prescriptive, with no visually sustained representations of gender ambiguity.

Gendered visual designs are reinforced through narrative roles

The visual gender differences identified in AI-generated mini-dramas were consistently reinforced through the assignment of narrative roles. Overall, male characters received more role assignments than female characters (42 vs. 35), indicating that male characters occupied a broader range of narrative functions across the sampled corpus.

Male characters were predominantly associated with agentic, conflict-oriented, and socially dominant roles. The most frequently assigned archetypes were Bystander ($n = 9$), Abuser ($n = 5$), Tormentor ($n = 5$), Cheater ($n = 4$), followed by Ally ($n = 3$) and Helper ($n = 3$). These roles positioned male characters as active participants in driving conflict, influencing other characters, or exercising authority within the narrative.

Female characters, by contrast, were more commonly portrayed in relationship-centred and emotionally driven roles. The most frequent archetypes were Affair Partner ($n = 4$) and Betrayed Partner ($n = 4$), followed by Bystander ($n = 3$), Victim of Abuse ($n = 2$), and Revenge Seeker ($n = 2$). Female characters also appeared as Cheaters ($n = 2$) and Tormentors ($n = 2$), but these roles occurred less frequently than among male characters. Overall, female roles were more strongly connected to interpersonal relationships, emotional conflict, betrayal, and victimisation than to positions of authority or independent agency.

Only two role assignments were associated with non-binary characters, both involving Manipulator and Schemer, while one Unclear character was coded as a Bystander. These findings indicate that gender diversity was almost entirely absent from the narrative role structure.

The role archetype analysis demonstrates that narrative functions reinforce the same gendered patterns observed in the visual design of characters. Male characters were more frequently positioned as active agents, authority figures, or sources of conflict, whereas female characters were primarily defined through relational, emotional, and victim-centred roles. Together, these findings suggest that AI-generated mini-dramas reproduce conventional gender stereotypes not only through visual appearance but also through the narrative functions assigned to characters.

Dialogue reinforces binary gender representation

Analysis of the dialogue transcripts revealed that gender was predominantly expressed through binary male and female lexical markers, with no explicit references to LGBTQ+ or non-binary identities detected across the corpus. A total of 57 male-related lexical indicators and 36 female-related lexical indicators were identified in the transcripts, demonstrating that male references occurred considerably more frequently than female references.

Male-related expressions were primarily conveyed through pronouns (e.g., *he*, *him*, *his*), family roles (e.g., *dad*, *father*, *son*, *hijo*), and relationship terms (e.g., *boyfriend*, *husband*). The most frequently occurring male-related terms included *hijo* ($n = 14$), *him* ($n = 10$), *he* ($n = 8$), and *dad* ($n = 6$). Female-related language similarly relied on pronouns and kinship terms, with the most common indicators being *her* ($n = 7$), *mamá* ($n = 6$), *girl* ($n = 5$), *she* ($n = 5$), and *sister* ($n = 5$). Importantly, no explicit LGBTQ+ or non-binary lexical indicators (e.g., *gay*, *lesbian*, *transgender*, *non-binary*, *they/them* as identity pronouns) were identified in any transcript. Instead, dialogue overwhelmingly relied on conventional male and female forms of address, family relationships, and gendered pronouns.

Gender Category	Most frequent lexical indicators	Frequency
Male	hijo	14
	him	10
	he	8
	dad	6
	el*	6
	boy	3
	sir	2
	señor	2
Female	her	7
	mamá	6
	girl	5
	she	5
	sister	5
	mom	3
	mama	2
	esposa	2
LGBTQ+	None detected	0
Non-binary	None detected	0

Table Frequency of gender-related lexical indicators identified in AI mini-drama transcripts

Overall, the transcript analysis demonstrates that AI-generated mini-dramas favour male over female gender representation in dialogue while reinforcing a binary understanding of gender. Male-associated language appeared substantially more frequently than female-associated language, and no explicit linguistic evidence of LGBTQ+ or non-binary identities was identified. These findings are consistent with the visual and narrative analyses, suggesting that AI-generated mini-dramas reproduce traditional binary gender norms across multiple modalities rather than presenting more diverse or inclusive gender representations.

AI constructs rather than merely reflects gender representation

A key finding of this study is that although fruit avatars are inherently gender-neutral, AI-generated mini-dramas systematically assign gendered meanings to them through visual appearance, clothing, body language, narrative roles, and dialogue. Rather than functioning as neutral fantasy characters, fruit avatars become symbolic representations of masculinity and femininity.

Viewed through a combined Gender Lens and Media Representation perspective, these findings suggest that AI-generated mini-dramas do not simply reflect existing gender norms but actively construct and reinforce conventional gender stereotypes by repeatedly associating masculinity with strength, authority, and agency, while linking femininity with attractiveness, emotional vulnerability, and relationship-oriented roles.

AI baby drama

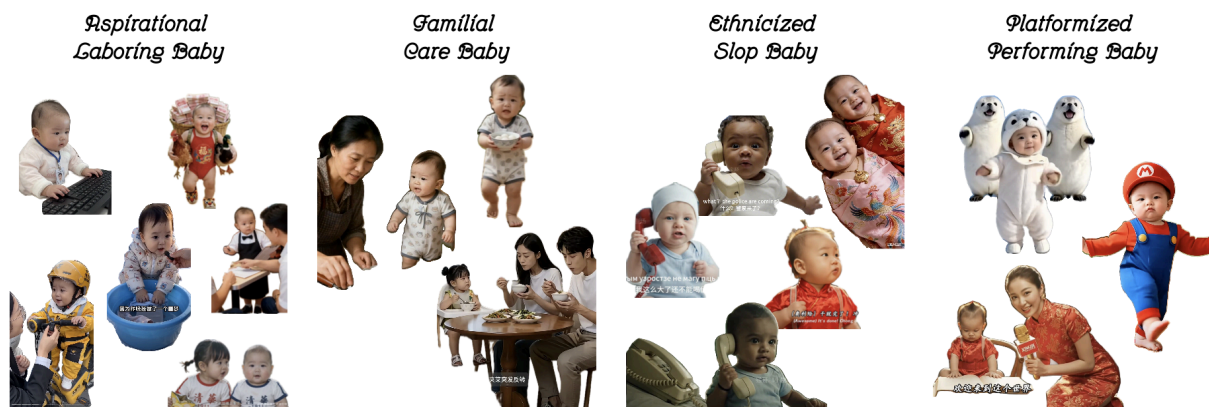


Fig 3. Illustration of the 4 types of AI baby trope

Aspirational Laboring Baby: The Infantilization of Social Pressure

The most salient feature in this category is the projection of adult roles onto AI babies. AI babies are repeatedly staged as small adults who enter recognisable social positions and conversations. For instance, they speak in the tone of mature adults, expressing concerns about wages, housing prices, and everyday survival. They become white-collar workers who are complained about by customers for not smiling during service. They also become exhausted food delivery riders who deliver orders late at night and eat instant noodles alone afterwards, as well as cashiers responding to workplace

demands, or students worrying about examinations and future success. Through role-playing between adults and AI babies, the baby becomes a speaking and acting subject who performs adult anxieties in an infantile form.

AI baby videos illustrate infant bodies as food delivery riders, cashiers, white-collar workers, and related social occupations, translating adult living pressure into cute and comic baby scenarios. In these videos, babies are often made to work, study, compete, earn money, purchase a house, or speak in the tone of exhausted adults, turning productivity and achievement into playful performances. This displacement makes social pressure affectively bearable because it is filtered through cuteness and humour (Ngai, 2012), yet it also suggests that aspiration, competition, and self-optimization are imagined as beginning before childhood properly begins (Kipnis, 2011). Additionally, the synthetic baby in these videos embodies a fantasy of productivity without biological maturity, becoming a figure through which Douyin visualizes and transforms social labouring as an affective attachment to infancy (Duffy, 2017).

Familial Care Baby: Re-staging Domestic Intimacy

AI baby videos also place infant bodies within domestic scenes of care, intimacy, and intergenerational attachment, such as being fed, comforted, accompanied, taught, or watched over by parents and grandparents. In the collected clips, the baby functions less as an autonomous individual than as a relational figure through which family affection and everyday care are made visible. In particular, these videos can be read as a platform-mediated translation of the Chinese cultural value of filial piety, where intergenerational care, emotional obligation, and family attachment are re-performed through Douyin's short-video aesthetics and AI-generated infant bodies (Yan, 2021). The hyper-realistic yet artificial infant intensifies the affective value of the child: it is not an actual family member, but it performs the sentimental position of the "priceless child" within the family imagination.

Meanwhile, because the baby is AI-generated, care becomes partially detached from biological kinship and redistributed across human viewers, platform interfaces, synthetic bodies, and algorithmic circulation. This finding shows that AI baby videos is not a simple expression of family love, but also re-stage family as a digitally mediated practice of care and display, where domestic intimacy becomes cute, shareable, and platform-ready (Blum-Ross and Livingstone, 2017).

Ethnicized Slop Baby: Cross-Cultural Templates of Infant Bodies

AI-generated babies are presented through cross-cultural comparison: "Chinese boy," "Russian boy," "American boy," "Indian boy," and other nationally or ethnically marked figures. The videos often repeat the same scenario across different cultural labels. For example, a baby may lie in a hospital bed, hold a telephone, or respond to a comic

situation, while the skin tone, facial features, language, subtitles, costume, and setting change from one version to another. In other examples, Chineseness is marked through red or pink embroidered clothing, dragon and phoenix motifs, gold pendants, festive fabrics, and auspicious visual symbols. The baby body becomes a flexible template through which ethnic and national identity is made instantly visible.

Analytically, these videos show how AI slop turns ethnicity into a repeatable visual formula. Cross-cultural difference is not represented through history, social context, or lived experience, but through quickly recognizable signs: labels, skin tone, accent, costume, background, and symbolic decoration. Because the figure is a baby, this ethnic typing appears cute, innocent, and playful. Yet this cuteness masks a classificatory logic: AI-generated infant bodies are sorted into simplified and comparable cultural types. The Ethnicized Slop Baby therefore, becomes a soft vehicle for platformized ethnic imagination, making global difference adorable, standardized, and easy to circulate.

Platformized Performing Baby: The AI-mediated Uncanny Cuteness

In this group of videos, AI-generated babies are shown singing and dancing to songs, sounds, and choreography that are currently popular across short-video platforms. Rather than performing generic children's songs, these baby-like figures reproduce Gen Z platform culture: viral music, dance challenges, meme gestures, exaggerated facial expressions, and influencer-style performance. The infant body is therefore placed inside a youth-oriented entertainment format, making the baby appear playful, funny, and immediately recognizable to platform audiences.

These videos reveal a tension between cuteness and uncanniness. Drawing on Sianne Ngai's theory of cuteness, the AI baby is made affectively appealing through smallness, softness, vulnerability, and apparent harmlessness. Cuteness here works as an aesthetic of emotional attachment: the infant body invites viewers to care for it, laugh at it, possess it, and circulate it as an adorable object. Yet this cuteness is also unstable. Because the baby sings and dances to trending Gen Z songs with highly coordinated movements, exaggerated facial expressions, and adult-like rhythm, the generated body exceeds the expected limits of infant embodiment. This produces an uncanny effect: the baby appears almost familiar and lovable, but its AI-mediated voice, movement, and facial animation reveal a nonhuman artificiality. Following Mori's "uncanny valley," the discomfort emerges precisely when a human-like figure becomes close to lifelike but remains perceptibly not-quite-human. In this sense, the platformized performing baby is not simply cute; it is a synthetic affective body in which cuteness and uncanniness are co-produced by AI animation, platform music, and short-video performance.

6. Discussion

AI baby drama

The four categories identified in this study — Aspirational Laboring Baby, Familial Care Baby, Ethnicized Slop Baby, and Platformized Performing Baby — collectively reveal that cuteness in AI baby videos is not an incidental aesthetic but a structural affective technology. Across all four types, the infant body functions as a flexible digital figure that absorbs and re-presents social content — labour pressure, family obligation, ethnic identity, and platform entertainment — in forms that are emotionally accessible, algorithmically shareable, and politically unmarked. Drawing on Ngai's (2012) theorization of cuteness as an aesthetic of power asymmetry, these videos demonstrate how the smallness and apparent vulnerability of the baby body makes otherwise heavy social content — precarious work, demographic anxiety, ethnic classification — bearable and even enjoyable. The synthetic infant does not simply represent these realities; it transforms them into objects of affective attachment and viral circulation.

At the same time, this study reveals an internal tension running through AI baby content: the simultaneous production of cuteness and uncanniness. While the Aspirational Laboring Baby and Familial Care Baby primarily operate through cuteness — making pressure and care feel intimate and warm — the Ethnicized Slop Baby and Platformized Performing Baby expose the limits of this affective regime. In the former, the repetitive visual templating of ethnic difference reveals a classificatory logic that cuteness merely disguises; in the latter, the AI baby's adult-like movement and vocal performance push the figure toward Mori's (1970) uncanny valley, where hyperrealism becomes estrangement. Together, these findings suggest that AI baby videos occupy an unstable affective position on Douyin: they are simultaneously designed for emotional resonance and structurally prone to generating discomfort, precisely because AI generation pushes infant bodies beyond the expected limits of human embodiment.

7. Conclusion

Discuss and interpret the implications of your findings and make recommendations for future research and application, be it societal, academic or technical (or some combination).

To conclude the study of AI Baby drama, the pilot study has examined AI baby videos on Douyin as a platform-specific form of synthetic affective media,

analysing how AI-generated infant bodies are made to perform adult labour, domestic care, ethnic identity, and viral entertainment. Through four empirically grounded categories, the analysis shows that the AI baby is not merely a cute digital novelty but a culturally loaded figure through which Douyin mediates and transforms the social life of its users. Cuteness, in this context, functions as an affective technology: it renders labour precarity comic, family obligation sentimental, ethnic difference adorable, and human-machine boundaries playful. In doing so, AI baby videos participate in the broader platformization of everyday feeling, where complex social and cultural tensions are absorbed into short, shareable, emotionally engaging content.

This study contributes to emerging scholarship on synthetic media, platform affect, and AI-generated content by demonstrating that the cultural significance of AI baby videos cannot be reduced to their technical novelty or visual appeal. As AI generation becomes increasingly embedded in short-video platforms, the infant body emerges as a particularly potent site for affective and ideological work — one that condenses anxieties about labour, kinship, race, and humanness into an apparently innocent form. Future research might extend this analysis by examining audience reception and comment cultures, comparing Douyin with TikTok to assess whether these affective formations are China-specific or reflect broader global trends in AI-generated content, or interrogating the political economy of the accounts and tools that produce this material at scale.

8. References

Blum-Ross, A. and Livingstone, S. (2017) "'Sharenting,' parent blogging, and the boundaries of the digital self", *Popular Communication*, 15(2), pp. 110–125.

Duffy, B. E. (2017) *(Not) Getting Paid to Do What You Love: Gender, Social Media, and Aspirational Work*. New Haven: Yale University Press.

Kipnis, A. B. (2011) *Governing Educational Desire: Culture, Politics, and Schooling in China*. Chicago: University of Chicago Press.

Mori, M. (1970) 'The uncanny valley', *Energy*, 7(4), pp. 33–35. [Trans. K. F. MacDorman and N. Kageki, *IEEE Robotics & Automation Magazine*, 19(2), 2012, pp. 98–100.]

Ngai, S. (2012) *Our Aesthetic Categories: Zany, Cute, Interesting*. Cambridge, MA: Harvard University Press.

Yan, Y. (ed.) (2021) *Chinese Families Upside Down: Intergenerational Dynamics and Neo-Familism in the Early 21st Century*. Leiden and Boston: Brill.