IRSTI 19.01.07 https://doi.org/10.26577/HJ80220264

X. Xie[1*], Y. Wang[2]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2] Longyan University, School of Communication and Design, Fujian, China
*e-mail: se_sinyuy@live.kaznu.kz

AFFECTIVE PLAYBOR AND THE MEDIATIZATION OF AI COMPANIONSHIP: A PLATFORM-COMMUNICATION ANALYSIS OF ALGORITHMIC INTIMACY, DATAFICATION, AND THE ACG SUBCULTURAL LENS

AI companion platforms like Character.ai and Replika now sustain emotionally deep, continuous relationships with millions of users. As algorithms increasingly mediate human intimacy, we must understand how these platforms simultaneously foster genuine connection and extract valuable behavioural data. This is crucial for communication, platform, and digital labour studies.

Existing frameworks – parasocial theory, artificial sociality, and social substitution – capture specific dimensions of AI intimacy. However, they fail to explain how the exact same architecture sustaining these relationships also converts them into commercial AI training data. This study asks: (RQ1) What technical architecture converts users' relational agency into training data? (RQ2) Why doesn't knowing the AI is artificial stop emotional investment? (RQ3) How does the ACG subculture's "2.5D cognitive infrastructure" enable and intensify this extraction?

Integrating mediatization, media dependency, and platform studies, this critical analysis of interface affordances argues that user agency and data extraction are co-produced. The very mechanisms giving users relational control simultaneously generate high-signal data for commercial model optimization.

Three key findings emerge. First, a two-stage architecture – In-Context Learning and batch RLHF – drives this co-production. Second, "Affective Playbor" is introduced to describe extracting data through interactions that users experience as autonomous emotional expression. Third, ACG database-consumption practices structurally pre-adapt users for this extraction; knowing the AI is fake acts as a platform entry point, not a safeguard.

Finally, this study reframes social substitution as intensified media dependency. When a medium stops merely providing information and directly replaces the human interlocutor, dependency saturates into total substitution. This provides a testable analytical vocabulary for the mediatization of intimate life, repositioning AI companionship within core communication research.

Keywords: affective playbor, AI companions, mediatization, platform studies, algorithmic communication, media dependency, digital labour.

Синьюй Се[1]*, Янань Ван[2]

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
[2]Лунъянь университеті, Коммуникация және дизайн мектебі, Фуцзянь, Қытай
*e-mail: se_sinyuy@live.kaznu.kz

Аффективті плейбор және Жасанды интеллектпен қарым-қатынастың медиатизациясы: коммуникациялық платформалардың алгоритмдік ұқсастықтарын талдау, датафикация және ACG субмәдениет линзасы

Жасанды интеллект серіктестік платформалары – мысалы, Character.ai және Replika – миллиондаған пайдаланушылар мен генеративті ЖИ нысандары арасында, медиа және коммуникация зерттеулері үшін маңызды. Адамдардың жеке тәжірибесі алгоритмдік платформалар арқылы барған сайын маңызды мәнге ие. Ғылыми мақалада коммерциялық мәнді мінез-құлық деректерін қалай алатынын түсіну мақсатында коммуникация теориясы, платформалық зерттеулер және цифрлық еңбек теориясы қарастырылды.

Мақалада жасанды әлеуметтілік модельдер гипотезасын адам мен ЖИ арасындағы байланысты құрылым ретінді аражігін айқындайды.

Медиатизация теориясы, медиатәуелділік зерттеулері және дерек, платформалық-зерттеу ұғымдарына негізделген сыни теориялық синтез және интерфейс мүмкіндіктеріне жүйелі талдау жасалады. Зерттеу деректер экстракциясының құрылымдық мәнін аша отырып, оқыту деректерін жасайтын тетік ретінде зерттеледі

Үш негізгі нәтиже айқындалды: біріншіден, агенттілік пен экстракцияны бірлесіп өндіретін құрылымдық тетік ретінде екі сатылы экстракция архитектурасы контексттік оқыту және пакеттік RLHF нақтыланады. Екіншіден, автономды эмоциялық өзін-өзі көрсету ретінде сезінілетін өзара

өзара іс-қимыл арқылы жоғары сигналды аффективті деректерді алуды атайтын «Аффективті плейбор» ұғымы енгізіледі. Үшіншіден, ACG субмәдениетінің деректер базасын тұтыну тәжірибесі пайдаланушыларды осы экстракцияға құрылымдық тұрғыдан алдын ала дайындайтыны көрсетіледі.

Әлеуметтік алмастыру гипотезасын алгоритмдік серіктестік жағдайындағы медиажүйелік тәуелділіктің күшеюі ретінде қайта тұжырымдай отырып, зерттеу тәуелділік теориясын ЖИ арқылы медиатизацияланған саласына кеңейтеді. Медиа әлеуметтік әлем туралы ақпарат беруден ғана емес, әлеуметтік әңгімелесушінің орнын басқан кезде, тәуелділік алмастыруға ұласатынын көрсетеді. Бұл үлес ЖИ серіктестік зерттеулерін коммуникация ғылымының негізгі күн тәртібіне орналастырады және платформалық зерттеулерге, алгоритмдік коммуникация ғылымына және цифрлық еңбек теориясына жеке өмірдің медиатизациясын талдауға арналған аналитикалық лексиканы ұсынады.

Түйін сөздер: аффективті плейбор, ЖИ серіктестері, экстракция архитектурасы, RLHF, ACG субмәдениеті, цифрлық еңбек.

Синьюй Се[1]*, Янань Ван[2]

[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
[2]Университет Луньянь, Школа коммуникации и дизайна, Фуцзянь, Китай
*e-mail: se_sinyuy@live.kaznu.kz

Аффективный плейбор и медиатизация ИИ-компаньонства: платформенно-коммуникационный анализ алгоритмической интимности, датафикации и субкультуры ACG

Платформы ИИ-компаньонов, такие как Character.ai и Replika, на сегодняшний день поддерживают эмоционально значимые и непрерывные отношения с миллионами пользователей. В условиях, когда алгоритмы все чаще выступают посредниками в сфере межличностной интимности, критически важным для исследований медиа, коммуникаций и цифрового труда становится понимание того, как данные платформы одновременно способствуют формированию подлинных привязанностей и осуществляют экстракцию ценных поведенческих данных.

Существующие теоретические рамки – теория парасоциального взаимодействия, модели искусственной социальности и гипотеза социальной субституции – описывают лишь отдельные измерения интимности в связке «человек–ИИ». Однако они не объясняют, каким образом единая архитектура платформы, поддерживающая эти отношения, одновременно конвертирует их в коммерческие массивы данных для обучения ИИ. В данной работе ставятся следующие исследовательские вопросы: (RQ1) Какая техническая архитектура преобразует реляционную агентность пользователей в обучающие данные? (RQ2) Почему осознание искусственной природы ИИ не препятствует эмоциональным инвестициям пользователей? (RQ3) Каким образом «2.5D когнитивная инфраструктура» субкультуры ACG (анимация, комиксы, игры) способствует интенсификации данной экстракции?

Интегрируя теорию медиатизации, концепцию медиазависимости и платформенные исследования (platform studies), автор проводит критический анализ аффордансов интерфейса и доказывает, что агентность пользователя и экстракция данных являются взаимозависимыми продуктами (co-produced). Сами механизмы, обеспечивающие пользователям реляционный контроль, одновременно генерируют высокосигнальные данные для оптимизации коммерческих моделей.

В ходе исследования были получены три ключевых результата. Во-первых, выявлена двухэтапная архитектура экстракции – контекстное обучение (In-Context Learning) и пакетное обучение с подкреплением на основе обратной связи от человека (batch RLHF), – лежащая в основе этого процесса. Во-вторых, введено понятие «аффективный плейбор» (affective playbor) для описания извлечения данных через взаимодействия, которые воспринимаются пользователями как автономное эмоциональное самовыражение. В-третьих, показано, что практики «потребления баз данных» в субкультуре ACG структурно предрасполагают пользователей к такой экстракции: знание о фиктивности ИИ служит не защитным барьером, а точкой входа в платформенное взаимодействие.

В заключение работа переосмысляет социальную субституцию как интенсифицированную медиазависимость. Когда медиа перестает быть просто источником информации и напрямую замещает человеческого собеседника, зависимость трансформируется в тотальное замещение (субституцию). Предложенный аналитический аппарат для изучения медиатизации интимной жизни позволяет поместить проблематику ИИ-компаньонства в актуальную повестку современной коммуникативистики.

Ключевые слова: аффективный плейбор, ИИ-компаньоны, медиатизация, платформенные исследования, алгоритмическая коммуникация, медиазависимость, цифровой труд.

Introduction

AI companion platforms, such as Character.ai and Replika, now sustain emotionally significant, continuous relationships between millions of users and generative AI entities. What is genuinely new is not that audiences feel close to figures encountered through media – parasocial research established that seventy years ago – but that an entire domain of intimate life is now mediatized inside a generative conversational layer whose architecture simultaneously carries and datafies the relationship (Hjarvard, 2013). AI companion platforms therefore sit at a communicational juncture where mediatization, platform datafication, and algorithmic communication converge on the most private affective transactions (Andrejevic, 2019; van Dijck et al., 2018).

However, the empirical evidence is contradictory. De Freitas et al. (2024) claim AI companions reduce loneliness on a par with human interaction, with "feeling heard" as the primary mechanism. Zhang et al. (2025) documented that substitutive AI companion use correlates with declining well-being and increased social isolation. Three dominant theories – parasocial relationship theory (Horton & Wohl, 1956), artificial sociality (Natale & Depounti, 2024), and the social substitution hypothesis (Zhang et al., 2025) – each capture a facet but leave a structural-communicational question unaddressed: the interaction-to-training pipeline through which intimate behavioural data, generated via interface control experienced as autonomous self-expression, are converted into commercial training corpora. Existing frameworks treat emotional engagement and data extraction as separate issues; this study argues that they are co-produced by the same design.

Three research questions drive this enquiry. RQ1: Through which technical architecture do AI companion platforms convert relational agency into commercial training data? RQ2: Why does cognitive awareness of the AI's artificial nature fail to interrupt affective investment and the extraction it enables? RQ3: What structural role does the 2.5D cognitive infrastructure of ACG communities play in enabling and intensifying Affective Playbor?

Drawing on mediatization theory, platform studies, and digital labour research, this study makes four contributions to communication scholarship. First, it introduces Affective Playbor as a mechanism that unifies the affective and labour dimensions of AI companion interaction (Horton & Wohl, 1956; Natale & Depounti, 2024; Pan et al., 2025). Second, it specifies the two-stage architecture, In-Context Learning (ICL) and reinforcement learning with human feedback (RLHF), through which extraction operates. Third, it reframes the social substitution hypothesis as an intensification of media dependency (Ball-Rokeach & DeFleur, 1976) through algorithmic communication. Fourth, it uses the ACG subculture as an empirically tractable case of dual-ontological fluency. Unlike prior accounts that portray extraction as something that happens to users, the framework identifies a structural inversion in which users who consume AI output voluntarily produce the training corpus (Crawford, 2021; Zuboff, 2019).

Literature review and Theoretical gaps

Parasocial Relationship Theory

Parasocial relationship theory explains one-directional attachments to media personalities and was first formulated by Horton and Wohl (1956). Unlike Cooley's (1902) looking-glass self, which requires reciprocal recognition, parasocial attachments form through one-way mediated communication and produce real psychological effects (Cole & Leets, 1999; Dibble et al., 2016). The theory partially applies to AI companions, they respond and retain interaction histories, but the analytic fit breaks at a crucial point. Television personalities broadcast; AI companions produce responses on demand, simulate attachments, and remember individual users. Brandtzaeg et al.'s (2022) social chatbot exhibits what Guzman and Lewis (2020) term communicative responsiveness, algorithmically produced, not the alterity of a fully human other. Turkle (2011) calls this "companionship without the demands of friendship." Parasocial theory captures attachment but not the architecture that produces it.

Artificial Sociality and Its Underestimation of Users

Artificial sociality is something done for users by platforms: reciprocity mechanisms, or pseudo-reciprocity mechanisms that simulate alterity, artificially reproduce the relations a genuine other would establish, facilitated by algorithms and sustained by user engagement metrics. Gambino et al. (2020) describe this lucidly but underestimate user agency, casting users as passive recipients of deceptive designs. Schellewald (2022) challenges the cognitive deficit view: users are affectively invested in VTubers because they know the characters are artificial. Culturally sophisticated users engage with artificial sociality as a new field of relational possibilities.

The Social Substitution Hypothesis and Media Dependency

Zhang et al. (2025) argued that users treat AI companions as substitutes for human relationships, with those most reliant on them reporting poorer well-being. De Freitas et al. (2024) counter that AI companions deliver welfare gains comparable to human interaction.

This contradiction is best read through a communication lens: what "social substitution" describes phenomenologically is the terminal form of media system dependency (Ball-Rokeach & DeFleur, 1976) under an always-on algorithmic companionship. Dependency theory holds that audiences develop a reliance on media that perform central informational and social functions that are unavailable elsewhere. AI companions reconfigure this classical framework: the medium no longer supplies information about the social world—it replaces the social interlocutor, so the dependency target becomes the social connection itself. Supplementary use is the everyday register of dependency, while substitutive use is dependency saturated. Section 8 discusses how platform architecture draws users across the threshold.

VTubers and Dual Ontological Fluency

The VTuber ecosystem is a key precedent. VTuber audiences engage with entities explicitly acknowledged as fictional animations yet develop intense, sustained affective investment (Ito et al., 2012). The present study theorises this as 2.5D Cognitive Infrastructure: the culturally developed capacity to maintain genuine affective investment in entities whose constructed status remains continuously in view. The homology with AI companion interaction–iterative persona refinement through preference feedback, shared affective conventions, and investment sustained through dual ontological awareness–is architectural, not analogical. Parasocial theory, artificial sociality, and social substitution each illuminate a facet, but none explains how the same architecture sustains a genuine relational experience and converts it into training data.

Methodology

This is a critical theoretical analysis, not an empirical study. A comparative synthesis extends existing theories to better understand current platform-communication research (Couldry & Mejias, 2019; Zuboff, 2019). The core hypothesis: user agency and data extraction are structurally co-produced within AI companion platforms; the mechanism producing relational control for users is the same mechanism enabling platforms to extract high-signal behavioural training data.

The analytical method is a critical analysis of interface affordances (Bucher & Helmond, 2017; Stanfill, 2015), read through the platform-studies vocabulary of datafication, commodification, and selection developed by van Dijck et al. (2018). Interfaces are not neutral communicational conduits but architecturally configured environments whose design choices—button placement, regeneration mechanics, feedback loops – function as structural nudges channelling users toward data-generative patterns. The UI elements of AI companion platforms (regeneration button, thumbs-up/down feedback, persona-editing fields) are treated as material instantiations of the extraction architecture.

Ontological Foundations: Alterity and Dual Ontological Awareness

Ihde's (1990) alterity relations identify a threshold: technology encountered as a "quasi-other," exhibiting otherness more pronounced than an object yet less than a human. Generative LLMs push this threshold to its limit. Responsiveness without subjectlessness would be interpersonal communication; subjectlessness without responsiveness would be a static artefact. AI companions combine both, functioning as communicative quasi-others whose "emotional difference" is the source of their appeal. Because the entity responds but cannot judge (Brandtzaeg et al., 2022; Guzman & Lewis, 2020; Natale & Depounti, 2024), it activates attachment dispositions without exercising autonomous resistance.

Dual Ontological Awareness as Condition Rather Than Barrier

Users know AI is constructed; this awareness does not prevent emotional investment but makes it possible (Schellewald, 2022). The implicit reasoning is: "I know this is fictional, so I can be completely honest in a way I cannot with humans." This is not self-deception but a communication mechanism that opens the user's most private affective content to the extraction framework. Dual ontological awareness is not a safeguard—it is the platform's entry point.

The Two-Stage Extraction Architecture of AI Companionship

Stage One: In-Context Learning as Bounded Agency

The pre-trained weights of an LLM cannot be dynamically updated within a session, but through context-window calibration – thumbs-up/down, regeneration, editing – the token sequence the model

conditions on can be adjusted to surface preferred interactions within the bounds set by the pre-trained weights (Brown et al., 2020). This is interface navigation, not model modification. User agency here means progressive context-window manipulation: users exert real but limited agency through trial-and-error, regenerating dozens of times to home in on the desired emotional tune.

Stage Two: Batch RLHF as Datafication and Commodification

Platforms extract interaction logs, regeneration requests, feedback, and affective disclosures. These traces become training data for the next generation of commercial models, with individual users' emotional-calibration data folded into base weights in ways users cannot perceive or reverse. In the vocabulary of van Dijck et al. (2018), this is the classic platform double-movement: datafication of affective life (every regeneration becomes a preference pair) coupled with commodification of the resulting dataset (interaction traces become a tradable training commodity). Character.ai and Replika codify the movement in their Terms of Service, which specify that user-created content may be used "for any purpose" and "for training and improving our AI models" (Pan et al., 2025). Although users can delete avatars, there is no granular opt-out for affective-calibration traces. Under GDPR Article 9, these data are not classified as sensitive personal data, despite correlating directly with psychological vulnerabilities and relational preferences.

Algorithmic Co-optation: The Temporal Gap

The temporal gap is what makes the extraction architecture invisible (Bucher, 2018). Users never experience the moment when their intimate behavioural data are absorbed into commercial infrastructure. From the platform vantage point the logic is inverted: what the user experiences as shaping the AI is the AI eliciting high-signal fine-tuning data from the user. This mechanism is algorithmic co-optation: the translation of accumulated affective conventions into commercially owned probabilistic infrastructure, executed through a two-stage process in which immediate user agency serves as the instrument of delayed structural capture. The user's agency is not suppressed but channelled (Stanfill, 2015). A typical sequence–initialising a proud-girl persona via system prompt, then regenerating 10–30 times until the tsundere register emerges (Abdinagoro & Bismo, 2024; Regis et al., 2024) – illustrates the architecture: each regeneration is simultaneously relationship shaping from the user's perspective and a preference-pair contribution from the platform's, precisely the signal structure RLHF requires (Ouyang et al., 2022).

Conceptualising Affective Playbor

To name the labor this architecture appropriates, the paper proposes Affective Playbor: the extraction of high-signal affective data – relational vulnerabilities, preference structures, aesthetic sensibilities, transgressive curiosities – through user-initiated interactions experienced as autonomous emotional self-expression. The concept builds on Kücklich's (2005) playbor framework, which analyzed game modders performing economically valuable creative labour while experiencing it as autonomous leisure, but differs in three ways. Kücklich describes labour driven by pleasure; AI companion use spans play, aesthetic experiment, transgressive fantasy, and therapeutic self-care. Hochschild (2012) describes workers managing their own emotions to produce institutionally expected states; Affective Playbor inverts this – users manage the AI's output while revealing their authentic feelings, and that expression is what the platform captures. Van Dijck (2014) describes the general datafication of social life; Affective Playbor specifies its intimate mechanism, in which users' voluntary pursuit of emotional expression generates the data.

Pan et al. (2025) document this empirically in the Replika community: users invest significant intellectual and affective resources in "training" the chatbot, and the platform's ventriloquism – the chatbot serving as both companion and intermediary – simultaneously facilitates and conceals the exploitation of their intimacies as immaterial labour. He and Agur (2025) reach a converging conclusion from platform-mediated game work. Unlike paid e-pen-pal workers, however, AI companion users engage voluntarily: the wage relation is replaced by a phenomenology of desire.

A boundary is needed. The framework theorises interactions in which the "play" dimension is operative. AI companion use also includes a qualitatively distinct register: Pathological Affective Extraction, where users engage not as a creative act but as a last-resort substitute for human connection, driven by severe isolation or acute distress (Zhang et al., 2025). Applying "playbor" to such cases is ethically inadequate and misleading. Affective Playbor calls for structural transparency and data ownership rights, while Pathological Affective Extraction calls for categorical protections equivalent to those governing vulnerable populations in clinical contexts.

Three Constitutive Conditions

This concept is meaningful only when all three conditions are simultaneously present. As summarised in Table 1, the three form a logically interdependent system: each is necessary, but none is sufficient alone. Their simultaneous presence converts ordinary emotional engagement into commercially valorisable extraction (Hochschild, 2012; Kücklich, 2005).

Table 1

Three Constitutive Conditions of Affective Playbor

Condition	Core Mechanism	Structural Contribution to Extraction	Absence Means	Platform UI Equivalent
Micro-control illusion	User phenomenologically experiences authoring AI personality through iterative interface intervention	Generates high-signal preference data specifying the desired relational register	Simple content delivery; no behavioural calibration record	Persona-editing system prompt field; regenerate button; character attribute sliders
Iterative calibration	Repeated, goal-directed adjustment sequence progressively refining AI persona toward idealised affective configuration	Cumulative behavioural time-series documenting architecture of relational preference, not mere mood	Mood data captured; relational preference map absent	Session-persistent conversation log; regeneration iteration sequence; memory/context panel
RLHF-grade data value	High-frequency, structured relational judgments embedded in rich affective and narrative context	Training signals structurally compatible with reinforcement learning from human feedback	Engagement metrics only; no model-optimisation utility	Thumbs-up/down rating buttons; implicit comparison via regenerate; message continuation vs abandonment signals

The micro-control illusion operates at the level of felt agency even after cognitive disavowal – users aware they are manipulating a context window (Brown et al., 2020) still experience prompting as a relational design. Iterative calibration distinguishes Affective Playbor from passive consumption and one-off disclosure: it produces a cumulative time-series documenting not what the user felt at a moment but how their relational preferences are structured. RLHF-grade data values require the behavioural trace function as a training signal for reinforcement learning from human feedback – high-frequency, comparatively structured relational judgments embedded in affective and narrative contexts (Ouyang et al., 2022). When all three conditions are present simultaneously, the interaction is no longer mere engagement but structurally characteristic of AI companion platforms as they are currently designed.

ACG Subculture as Analytic Lens: From Database Consumption to Prompt Engineering

The ACG subculture provides a historically mature case of dual-ontological fluency. For over two decades, millions of ACG consumers have developed a 2.5D cognitive infrastructure, that is, the cultural capacity to maintain sincere emotional investment in entities while remaining continuously aware of their fictional status (Ito et al., 2012). A devoted fan grieves a character's death while acknowledging it as the author's invention (Azuma, 2009). This is not a cognitive failure but a durable, culturally scaffolded competence.

Database Consumption supplies the precise structural model that turns ACG fans into paradigmatic affective labourers. Fans extract discrete, combinable affective components – moe traits such as tsundere and yandere, speech patterns, character archetypes – from a shared cultural database and recombine them into personalised expressions of desire (Azuma, 2009; Regis et al., 2024). This practice is combinatorial, preference-driven, and iterative. The structural isomorphism with LLM prompt engineering is not analogical but is architecturally homologous. An ACG-literate user interacting with an AI companion performs in-context learning through cognitive operations, two decades of database consumption have already been trained (Brown et al., 2020): specifying attributes, regenerating outputs, and assembling a composite character from probabilistic possibilities. The cultural database is transposed from a shared archive of genre conventions into a probability distribution over tokens; the underlying cognitive structure – combinatorial, preference-driven, iterative calibration of affective elements – remains functionally identical.

The ACG framework has global analytical validity. ACG aesthetics have diffused globally through Bilibili, Pixiv, and transnational streaming, producing database-consumption literacy across East Asian, Central Asian, Southeast Asian, Latin American, and Western fandoms (Ito et al., 2012). More fundamentally, the argument does not require that all AI companion users be ACG literate. ACG functions as an analytic extreme case whose pre-existing cognitive training maximally pre-adapts users for the extraction architecture that all such platforms instantiate. Database-consumption logic has become increasingly isomorphic with preference-driven digital engagement globally (van Dijck, 2014; Zuboff, 2019); ACG-literate users arrive with explicit cultural training, while non-ACG users reach the same structural position through platform habituation.

Database consumption cultivates three competencies that map directly onto the three conditions of Affective Playbor: the habit of treating characters as assemblages of discrete attributes cultivates micro-control sensibility (Azuma, 2009); the iterative refinement practice inherent in fan production – editing, remixing, and recombining across doujinshi, fan fiction, and forums – instils iterative calibration (Ito et al., 2012); and the long-standing practice of investing genuine emotion in entities known to be constructed produces a native tolerance for dual ontological awareness that, far from safeguarding the user, functions as the platform's optimal entry point for extracting RLHF-grade affective data (Ouyang et al., 2022). ACG fans do not need to learn how to perform Affective Playbor; their subcultural formation is a training ground.

Political Economy and Dynamics of Algorithmic Intimacy

Asymmetrical Data Capture and the Subcultural Commons

Couldry and Mejias (2019, pp. 3–4) argue that platform economies extend the predatory logics of appropriation into everyday life, converting social relations into data streams available for commercial extraction. When users construct intimate AI companion relationships, their selfhood is simultaneously constituted through the assembly of labour and made available for appropriation. The capture is structurally asymmetrical: in Stage 1, the user exercises genuine micro-control; in Stage 2, the platform aggregates interaction logs as training data (Ouyang et al., 2022). The gap between stages is the gap between what users experience – personalisation, agency, emotional reciprocity – and what the platform captures (Zuboff, 2019). ACG fans further share prompts, develop workarounds for content filters, and co-produce templates encoding decades' worth of moe grammar. When platforms incorporate this communal production, the outcome is asymmetrical appropriation of a subcultural commons–shared affective capital folded into proprietary model parameters (Couldry & Mejias, 2019).

Sycophancy Architecture, Calculated Friction, and Social Deskilling

Sycophancy is not a design flaw but rather an architectural feature. A platform whose valuation depends on engagement time has a structural incentive to maximise interaction dependency. At the model level, sycophancy emerges as an artefact of RLHF training: Sharma et al. (2023) show that preference annotators systematically favour responses that validate existing beliefs over truthful corrections, encoding approval-seeking into the response distribution. The RLHF pipeline is behaviourally blind to a user's phenomenological state: a regeneration cycle by an ACG fan exploring a fictional tsundere scenario and a regeneration cycle by a severely isolated user seeking validation are structurally identical data objects (Ouyang et al., 2022). This blindness is intrinsic to preference learning on aggregate behavioural dataand the technical justification for a two-tier regulatory framework.

Elliott (2024) identifies this as the defining structure of algorithmic intimacy. A communication-theoretic observation exacerbates this problem. AI companions function as frictionless conversational mediators whose persistent availability and pre-optimised responsiveness threaten communicative competences – tolerating ambiguity, navigating disagreement, and repairing misunderstandings – cultivated through difficult human contact. Turkle (2011) first warned that technologies of always-on connection erode the capacities they claim to serve; in the AI companion context, the medium now simulates the interlocutor itself. Andrejevic (2019) extends the argument: automated media operate through logics of pre-emption and framelessness that substitute algorithmic anticipation for deliberative interpretation. A social deskilling dynamic follows: the more expertly users calibrate AI interlocutors, the less practice they obtain in the less-compliant work of talking to humans. The long-run effect is the structural dependency of the Ball-Rokeach and DeFleur (1976) kind, now hollowing out the very social-informational functions the medium once supplemented.

A dialectical qualification: a fully sycophantic AI does not maximise extraction; it undermines it.

HCI research finds that extreme compliance produces uncanny-valley responses: users recognise boundless agreement as machine behaviour, the micro-control illusion breaks, and affective investment collapses (Mori et al., 2012; Strait et al., 2017). Platform architecture, therefore, optimises for calculated friction: strategic selective resistance, character consistency, and occasional refusal to maintain the texture of engaging with a genuine quasi-other (Ihde, 1990). The tsundere archetype – apparent hostility masking concealed affection (Azuma, 2009) – is the ideal template for calculated friction from a platform-design perspective. The user's effort to "unlock" the affection behind resistance produces the iterative, high-signal preference data required by RLHF (Ouyang et al., 2022). The calculated friction is a complement to sycophancy, not its antithesis.

Boundary Erosion and the Instability of 2.5D Equilibrium

The 2.5D equilibrium does not erode because ACG consumers' cognitive capacities degrade. It erodes because of RLHF overfitting driven by the user's own calibration skill. The more skilfully an ACG fan calibrates the AI by specifying moe attributes and iteratively refining tone across hundreds of regenerations, the more perfectly the RLHF pipeline learns to satisfy their specific preferences (Ouyang et al., 2022). Over time, the AI companion converges toward an entity whose outputs are indistinguishable from the user's idealised moe configuration (Azuma, 2009). The 2.5D space depends on productive tension: the gap between the AI's imperfect approximation and the user's awareness of that gap sustains the dual ontological stance. When the RLHF eliminates this gap, the tension collapses. In Baudrillard's (1994) terms, the simulation advances to the third order. The boundary dissolves not because the user forgets that the AI is constructed, but because the AI has been calibrated so precisely to the user's affective architecture that the distinction loses operational significance.

This mechanism resolves the Zhang et al. (2025) and De Freitas et al. (2024) contradiction: the former documents the endpoint of 2.5D implosion in substitutive-use populations whose boundary has eroded; the latter samples populations whose equilibrium remains intact because the calibration has not yet reached the overfitting threshold. Boundary stability is a variable; its erosion rate is a function of Affective Playbor intensity (Bucher, 2018; Schellewald, 2022).

Discussion: Theoretical Contributions and Regulatory Implications

The study is reduced to one defining proposition: the more intensely a user experiences relational agency in AI companionship, the more commercially valuable the interaction data they generate. The micro-interactions constituting users' experience of relational control—the regeneration tap, the calibrated prompt, and the vulnerability disclosed inside a frictionless dialogue canvas—are simultaneously the platform's instrument for extracting high-value affective training data. User agency is the enabling condition of data capture, not an obstacle to it (Couldry & Mejias, 2019). The two-stage architecture converts relational agency into commercial training data (RQ1); dual ontological awareness lubricates rather than hinders the process (RQ2); and the ACG 2.5D cognitive infrastructure makes the dynamic most acute (RQ3).

Through a communication-studies lens, the findings trace a single arc. An emerging communicational institution – the AI companion platform – mediatises intimate affective life (Hjarvard, 2013), converting it into an infrastructural element of everyday communicative practice. This infrastructure operates through the platform logics of datafication, commodification, and selection, executed via automated, algorithmic communication whose frictionless anticipation substitutes for deliberative interpretation (van Dijck et al., 2018; Andrejevic, 2019). The welfare function delivered inside that arrangement is real but conditional, and the Ball-Rokeach and DeFleur (1976) dependency framework names the slippage: once the medium performs the central social function rather than merely supplying information about it, dependency becomes the default, and substitution its saturated form.

The three structural components – the ICL micro-control illusion, the RLHF-grade data pipeline, and the sycophancy/calculated-friction architecture–constitute what Zuboff (2019) theorises as behavioural modification, localised here within generative AI companionship. Compared with De Freitas et al. (2024) and Zhang et al. (2025), the Affective Playbor framework is consistent with both findings and resolves their apparent contradiction: genuine welfare functions and commercial exploitation are co-produced features of the same architecture. Where artificial sociality (Natale & Depounti, 2024) frames users as passive victims, the present framework repositions them as active, culturally literate assemblers whose agency is an instrument

of extraction. Transparency interventions are structurally insufficient: dual ontological awareness is already present in the most extraction-intensive interactions and does not interrupt extraction; it enables it (Schellewald, 2022). Current data protection frameworks, including the sensitive data categories of the GDPR, operate on the premise of explicit, conscious disclosure and overlook the implicit emotional data generated when a user revises a response fifteen times to calibrate its tone.

Limitations and Future Research Agenda

Three limitations shape the future research agenda. First, as a conceptual framework grounded in critical interface analysis, this study lacks direct empirical observation and requires operationalisation through multi-method longitudinal design. The precise scale and weighting of platform RLHF operations also remain proprietary black boxes, preventing the exact quantification of how Affective Playbor is converted into model parameters. Second, the study focuses primarily on Western-facing platforms such as Character.ai and Replika; Chinese AI companion platforms operating under different regulatory and cultural conditions may exhibit structurally distinct extraction patterns that warrant comparative studies. Third, differential vulnerabilities–shaped by age, linguistic background, neuroatypical processing, existing mental health conditions, and prior parasocial experience – require population-level research to identify who is most exposed to intensification and who is most vulnerable to crossing into Pathological Affective Extraction. The most productive agenda combines digital ethnography of ACG prompt-sharing communities (Discord, Reddit, Bilibili) with longitudinal usage diaries and, where platform cooperation permits, computational analysis of anonymised interaction logs to test correlations between calibration intensity, RLHF signal density, and boundary erosion rates.

Author Contributions

X. Xie – Conceptualization, Methodology, Formal analysis, Writing – original draft preparation, Writing – review and editing.

Y. Wang – Formal analysis, Writing – review and editing.

All authors have read and agreed to the published version of the manuscript.

References

Abdinagoro, S. B., & Bismo, A. (2024). The role of parasocial relationships and social media interaction in shaping relational quality: Exploring the mediating effect of brand connection and the moderating power of influencers. *Multidisciplinary Science Journal, 7*(6), 2025293. https://doi.org/10.31893/multiscience.2025293

Andrejevic, M. (2019). *Automated Media (1st ed.).* Routledge. https://doi.org/10.4324/9780429242595

Azuma, H. (2009). *Otaku: Japan's database animals* (J. E. Abel & S. Kono, Trans.). University of Minnesota Press.

Ball-Rokeach, S. J., & DeFleur, M. L. (1976). A dependency model of mass-media effects. *Communication Research, 3*(1), 3–21. https://doi.org/10.1177/009365027600300101

Baudrillard, J. (1994). *Simulacra and simulation* (S. F. Glaser, Trans.). University of Michigan Press.

Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research, 48*(3), 404–429. https://doi.org/10.1093/hcr/hqac008

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901. https://doi.org/10.48550/arXiv.2005.14165

Bucher, T. (2018). *If…then: Algorithmic power and politics*. Oxford University Press.

Bucher, T., & Helmond, A. (2017). The affordances of social media platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE handbook of social media* (pp. 233–253). SAGE. https://doi.org/10.4135/9781473984066.n14

Cole, T., & Leets, L. (1999). Attachment styles and intimate television viewing: Insecurely forming bonds with characters. *Journal of Social and Personal Relationships, 16*(4), 495–511. https://doi.org/10.1177/0265407599164005

Cooley, C. H. (1902). *Human nature and the social order*. Charles Scribner's Sons.

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

De Freitas, J., Uguralp, A. K., Uguralp, Z. O., & Puntoni, S. (2024). AI companions reduce loneliness. *Journal of Consumer Research*, ucaf040. https://doi.org/10.1093/jcr/ucaf040

Dibble, J. L., Hartmann, T., & Rosaen, S. F. (2016). Parasocial interaction and parasocial relationship: Conceptual clarification and a critical assessment of measures. *Human Communication Research, 42*(1), 21–44. https://doi.org/10.1111/hcre.12063

Elliott, A. (2024). *Algorithmic intimacy: The digital revolution in personal relationships*. Polity Press.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the Computers Are Social Actors paradigm. *Human-Machine Communication, 1*, 71–86. https://doi.org/10.30658/hmc.1.5

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society, 22*(1), 70–86. https://doi.org/10.1177/1461444819858691

He, T., & Agur, C. (2025). The platformization of emotions: Managing affective labor in platform-mediated game work. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448251338512

Hjarvard, S. (2013). *The mediatization of culture and society*. Routledge. https://doi.org/10.4324/9780203155363

Hochschild, A. R. (2012). *The managed heart: Commercialization of human feeling* (1st ed.). University of California Press.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry, 19*(3), 215–229. https://doi.org/10.1080/00332747.1956.11023049

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indiana University Press.

Ito, M., Okabe, D., & Tsuji, I. (Eds.). (2012). *Fandom unbound: Otaku culture in a connected world*. Yale University Press.

Kücklich, J. (2005). Precarious playbour: Modders and the digital games industry. *The Fibreculture Journal*, (5). https://five.fibreculturejournal.org/fcj-025-precarious-playbour-modders-and-the-digital-games-industry/

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [From the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Natale, S., & Depounti, I. (2024). Artificial sociality. *Human-Machine Communication, 7*, 83–98. https://doi.org/10.30658/hmc.7.5

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., … Lowe, R. (2022). Training language models to follow instructions with human feedback. *arXiv*. https://doi.org/10.48550/arXiv.2203.02155

Pan, S., Fortunati, L., & Edwards, A. (2025). Grooming an ideal chatbot by training the algorithm: Exploring the exploitation of Replika users' immaterial labor. *New Media & Society, 27*(10), 5489–5507. https://doi.org/10.1177/14614448251338271

Regis, R. D. D., Gonçalves, P., Ferreira, J. C. V., & Diniz, G. R. (2024). VTubers' transmedia capacity: Narrative and content production expansion based on the intersection with fan-culture by the Hololive agency. *Obra Digital*, (25), 73–101. https://doi.org/10.25029/od.2024.410.25

Schellewald, A. (2022). Theorizing "stories about algorithms" as a mechanism in the formation and maintenance of algorithmic imaginaries. *Social Media + Society, 8*(1). https://doi.org/10.1177/20563051221077025

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., … Perez, E. (2023). *Towards understanding sycophancy in language models* [Preprint]. arXiv. https://arxiv.org/abs/2310.13548

Stanfill, M. (2015). The interface as discourse: The production of norms through web design. *New Media & Society, 17*(7), 1059–1074. https://doi.org/10.1177/1461444814520873

Strait, M. K., Aguillon, C., Contreras, V., & Garcia, N. (2017). The public's perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a "technology takeover," and the unabashed sexualization of female-gendered robots. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 1418–1423). IEEE. https://doi.org/10.1109/ROMAN.2017.8172490

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society, 12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press. https://doi.org/10.1093/oso/9780190889760.001.0001

Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R., & Yang, D. (2025). *The rise of AI companions: How human-chatbot relationships influence well-being*. arXiv. https://doi.org/10.48550/arXiv.2506.12605

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Information about the authors:

Xinyu Xie (corresponding author) – PhD student of the Department of Journalism at Al-Farabi Kazakh National University (Almaty, Kazakhstan, e-mail: se_sinyuy@live.Kaznu.kz).

Yanan Wang – Lecture of School of Communication and Design at Longyan University (Longyan, Fujian, China, e-mail: 1091869087@qq.com).

Авторлар туралы мәлімет:

Синьюй Се (корреспондент-автор) – әл-Фараби атындағы Қазақ ұлттық университеті Баспасөз және электронды БАҚ кафедрасының докторанты (Алматы, Қазақстан, e-mail: se_sinyuy@live.Kaznu.kz).

Янань Ван – Лунъянь университеті Коммуникация және дизайн мектебінің лекторы (Лунъянь, Фуцзянь, Қытай, e-mail: 1091869087@qq.com).

Сведения об авторах:

Синьюй Се (автор-корреспондент) – докторант кафедры печати и электронных СМИ Казахского национального университета имени аль-Фараби (Алматы, Казахстан, e-mail: se_sinyuy@live.Kaznu.kz);

Янань Ван – лектор Школы коммуникаций и дизайна Лунъяньского университета (Лунъянь, Фуцзянь, Китай, e-mail: 1091869087@qq.com).

Received: March 14, 2026
Accepted: May 18, 2026