Selected Papers of #AoIR2025:
The 26th Annual Conference of the
Association of Internet Researchers
Niterói, Brazil / 15 – 18 Oct 2025

CULTURAL BAIT: KWAI'S COLD START ALGORITHM AND THE INSTRUMENTALIZATION OF BRAZILIAN CULTURE

Elias Bitencourt
State University of Bahia (UNEB)

Guilherme Bispo C. Santos.
State University of Bahia (UNEB)

Nusta Oviedo
State University of Bahia (UNEB)

Rayssa Keuri Pereira Batista
State University of Bahia (UNEB)

Cecilio Ricardo de Carvalho Bastos
State University of Bahia (UNEB)

Introduction

Short-video platforms have become central to digital cultural production, shaping how local cultures are represented and monetized (Liu, 2020; Lin & de Kloet, 2019). Existing research often uses Bucher's algorithmic imaginary to examine user perceptions of algorithmic agency or to show how algorithmic logic reorganizes visibility regimes (Bucher, 2012, 2016, 2017; Arriagada & Ibáñez, 2020; Geboers & Van De Wiele, 2020; Jurno & D'Andréa, 2017; Karhawi, 2024; Araujo & Karhawi, 2024; Strecker, 2024). Work on Kuaishou tracks visibility among "invisible classes" and rural labor, as well as broader cultural platformization (Liu, 2020; Tan et al., 2020; Xi, 2024; Zhou & Liu, 2021; Lin & de Kloet, 2019). Yet these studies largely reduce the algorithmic imaginary to user perceptions, overlooking infrastructural and computational constraints on cultural production.

A second strand uses audits and platform observability, treating rankings and recommendations as outputs to infer how platforms allocate visibility over time (Jokubauskaitė et al., 2023; Matamoros-Fernández et al., 2021; Rieder et al., 2018; Rieder & Hofmann, 2020). This work documents temporal dynamics and winner effects

Suggested Citation (APA): Bitencourt, E., Santos, G. B. C., Oviedo, N., Batista, R. K. P., & Bastos, C. R. C. (2025, October). *Cultural bait: Kwai's cold start algorithm and the instrumentalization of Brazilian culture*. Paper presented at AoIR2025: The 26th Annual Conference of the Association of Internet Researchers. Niterói, Brazil: AoIR. Retrieved from http://spir.aoir.org.

but, because it studies systems in operation, is less suited to explore how platforms construct cultural preferences for new users at first exposure.

This paper examines Kwai's cold-start phase, when the recommender operates without prior user data and shapes initial visibility by selecting and amplifying culturally recognizable content. Rather than user perceptions, collective productions, or audits of ranking and recommendation dynamics, we focus on the cultural imaginary assembled by recommendation algorithms at first exposure, when platforms begin to infer unknown communities. We ask which cultural assumptions the algorithm encodes and how it uses them to attract, engage, and retain new users. We propose cultural bait to name platforms' strategic curation, reduction, and amplification of culturally recognizable content as an epistemic and computational tool to capture the attention of unknown users.

Kwai Cold Start: Early Imaginaries Based on Computational Efficiency and Marketing Strategies

Since entering Brazil in 2019, Kwai has grown rapidly: Brazil now accounts for 80% of its foreign revenue, 50% of its non-Chinese daily users, and about 60 million active users (Flach, 2024). Targeting emerging economies and working-class audiences, Kwai invests 15% of its Brazilian budget to compensate 400,000 of 2.1 million creators by content performance, a monetization model that helps retain marginalized audiences overlooked by dominant platforms (Deck & Marasciulo, 2022).

This strategy extends to Kwai's recommendation algorithm as well. Unlike TikTok's real-time micro-interaction tracking with deep learning models, Kwai employs a cache-aware reinforcement learning (CARL) framework (Chen et al., 2024) that prioritizes preloaded cached content over dynamic personalization during cold start.

The cold start phase is relevant because, without personalized data, the platform relies on pre-assumed cultural cues and serves new users a default set of global and regional "top hits" that provide an entry point into the content ecosystem and a visibility hierarchy for circulation within optimized audience niches. Videos that perform well in this window are cached and amplified; those that don't are deprioritized and removed.

Relying on cached content during cold start enhances computational efficiency by reducing real-time inference and latency. This stabilizes recommendations in weaker infrastructures and reinforces cultural prioritization, amplifying content aligned with relevant engagement patterns rather than relying on individual preference modeling.

Study Design

We adopted an exploratory design with four simulations, balanced by gender and location, to assess invariance in early recommendations. Four anonymous users were simulated using a Python JSON-scraping script that accessed publicly available web client endpoints. Sessions were genuine cold starts (no cookies, logins, prior history, VPN, or IP masking), with no simulated viewing or interactions (no likes, follows, shares, comments, or manual skips), and pagination was used only to retrieve initial recommendation batches.

We collected ≈1,000 recommendations per user at random intervals in December 2024. For each recommended video, we captured the main thumbnail returned in the JSON payload as the pre-click visual; audio and captions were not analyzed. Users B and C used the same Wi-Fi to observe possible IP-level clustering.

We processed ≈ 4,000 thumbnails with Smart Imagery Mapping (SIM; Bitencourt, 2023), embedding them with a Vision Transformer (ViT) aligned with Kwai/Kuaishou's transformer-based multimodal architecture (QUARM; Luo et al., 2024), reducing embeddings via PCA and UMAP, and clustering with HDBSCAN, with hyperparameters tuned by grid search and internal cross-validation using Silhouette and Davies–Bouldin indices (Hotelling, 1933; McInnes et al., 2020; Campello et al., 2013; Rousseeuw, 1987; Davies & Bouldin, 1979).

To compare distributions across users, we used the Jensen–Shannon divergence (Lin, 1991) and the Chi-square test (Pearson, 1900). Clusters were classified as homogeneous transversal topics (present across all users with no significant variation, $p > 0.05$), non-homogeneous transversal topics (present across users but in varying proportions), or niche clusters (appearing only for specific profiles). Labels were assigned after qualitative rounds and author consensus. All data came from public JSON sources; no personally identifiable information was collected, and analyses used only aggregate artifacts such as post labels and image embeddings.

Findings

During the cold start phase, Jensen–Shannon distances among the four users' clusters are very low, indicating very similar topic distributions across users (Fig. 1).

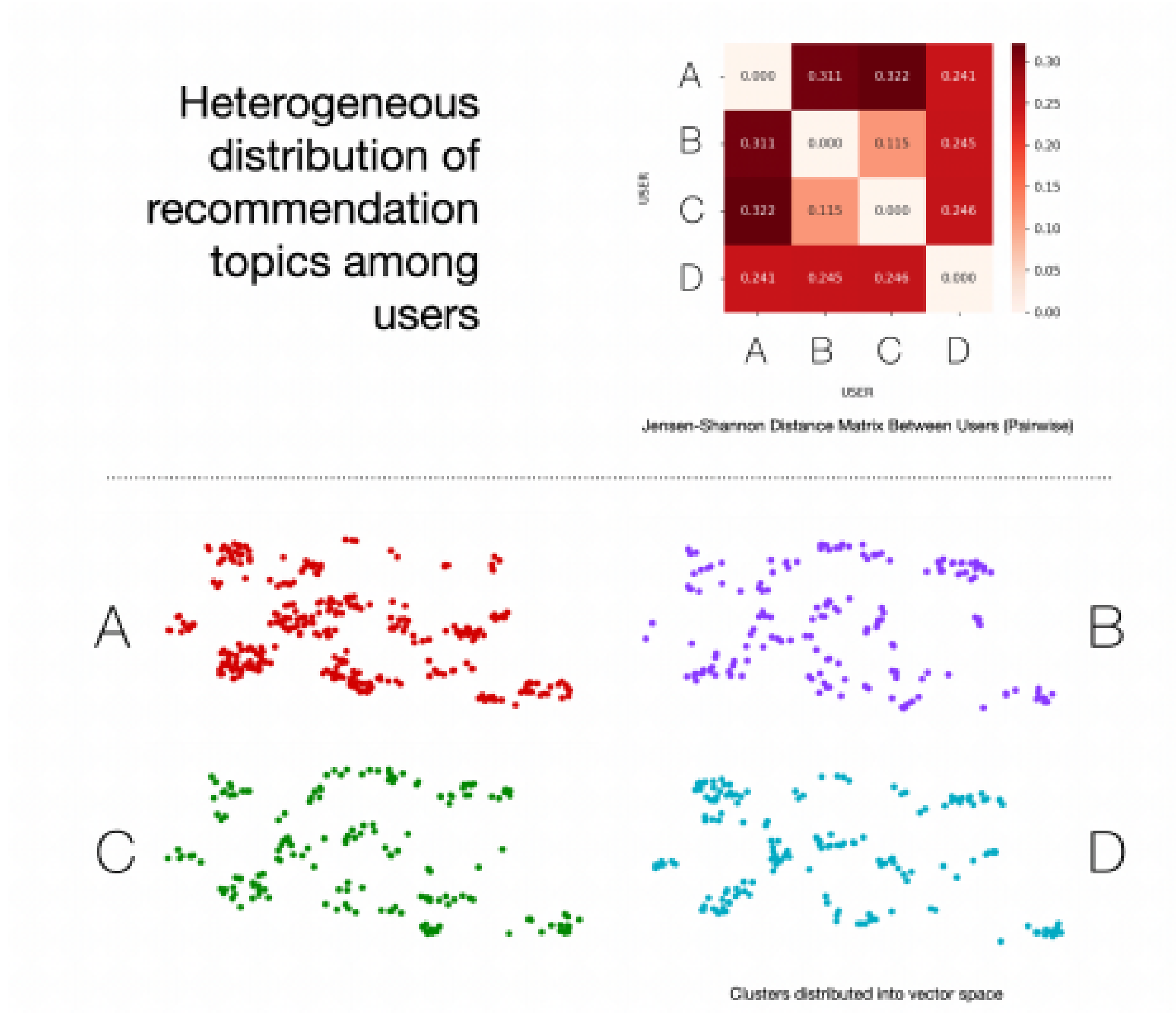

Fig. 1 Distribution of recommendation topics among users.

Transversal topics account for 96.8% of the recommended content. Among these, homogeneous transversal topics—which represent 56% of all clusters—comprise 54.62% of Kwai's cold start content, with Series/Telenovelas (12.65%), Football Culture (10.32%), Scripted Drama (7.84%), Suggestive Humor (6.88%), and Relationship Clips (5.23%) standing out (Fig. 2).

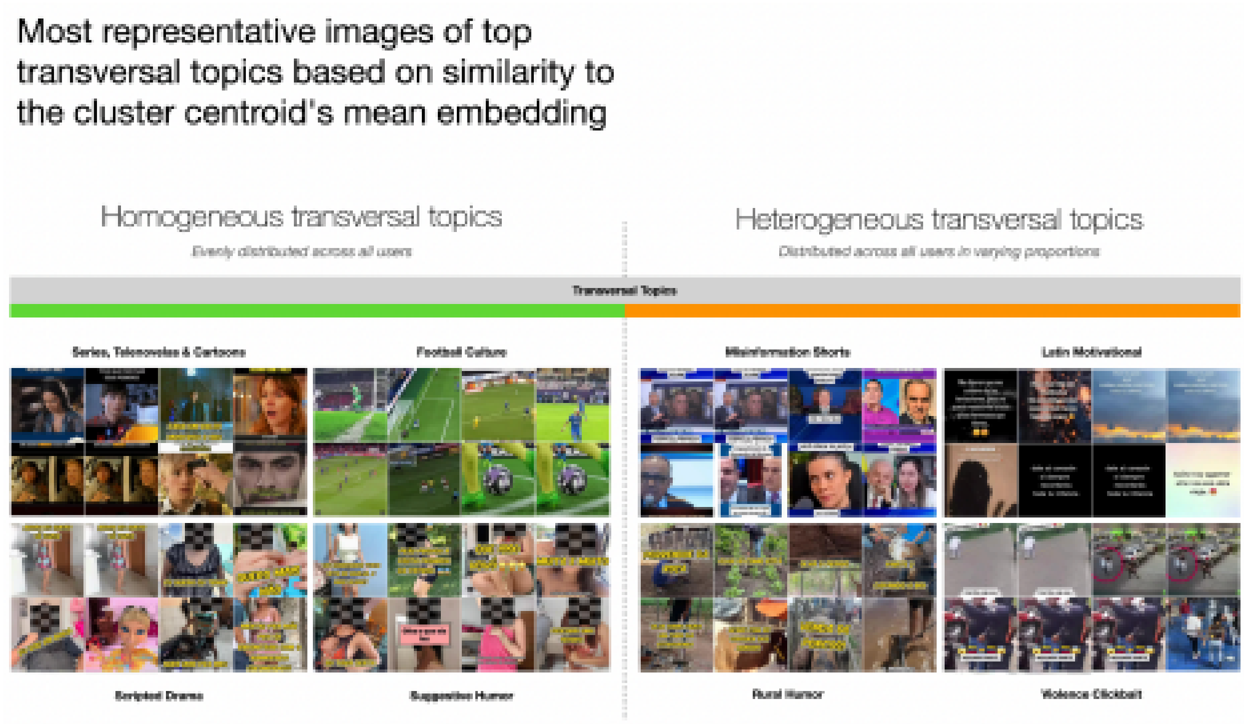

Fig. 2 Key recommended transversal topics.

Heterogeneous transversal topics—38% of all clusters—highlight Misinformation Content (9.49%), Latin Motivational (9.35%), Rural Humor (8.25%), and Violence ClickBaits (5.91%). Niche topics (6% of all clusters) focus solely on first-person shooter (FPS) games.

Homogeneous clusters mark the standard content base shared by all users, whereas heterogeneous clusters signal controlled adjustments that calibrate early exposure. Even with four test profiles, near-zero Jensen–Shannon distances indicate little between-user difference, suggesting these patterns are system-level rather than user-driven and signaling an algorithmic imaginary that precedes the construction of cultural patterns based on user-specific learning.

Results indicate that the cold start follows an exposure formula that primarily relies on a homogeneous content base aligned with international stereotypes of Brazilian culture (telenovelas, football, sensualization), combined with short dramas imported from China and adapted for Brazil as a user-retention strategy[1], while more controversial heterogeneous content is used to test user acceptability.

Cultural Bait: Algorithmic Speculation and the Scalable Instrumentalization of Culture

[1] see: https://www.terra.com.br/byte/mininovelas-sao-aposta-do-kwai-para-prender-usuarios,5acbf6fef899acd1983e0d370cc6e2504xb3wyjf.html

Our analysis of Kwai's Brazilian cold-start strategy extends work on algorithmic cultural production (Liu, 2020; Zhou & Liu, 2021; Lin & de Kloet, 2019) by foregrounding what we call “Cultural Bait”: modular and adaptive dispositive that leverage caricatured cultural tropes (e.g., football, telenovelas) as economical, computationally lightweight heuristics; content solid enough to resonate immediately yet flexible enough to be repurposed across markets.

The “Cultural Bait” exposes the logic of Kwai’s algorithmic epistemologies, which, from the first user encounter, intertwine with technical efficiency, market-driven choices, and geopolitical strategy, positioning culture as a resource for infrastructural and economic expansion. It reduces cultural expressions to latent, infrastructurally optimized modules that balance familiarity and modularity to know unknown communities and commodify engagement. Cold start is thus neither neutral nor random but embedded in Kwai’s socio-technical logic, where bandwidth constraints and low-end devices shape how culture is decomposed, repackaged, and redistributed.

Although Kwai’s algorithm is not generalizable across platforms, as each encodes a distinct political algebra and produces specific cultural distortions, examining its CARL framework illuminates an epistemological shift in which algebra, latent spaces, and discrete representations codify cultural meaning. Media studies often retreat into critiquing algorithms as black boxes or proprietary systems. Even if they are institutionally and operationally opaque, their mathematics and logic remain universal, readable, and open to interdisciplinary engagement.

References

Araujo, W. F., & Karhawi, I. (2024). “Todo mundo pode ser famoso com o algoritmo do TikTok”. In Imaginários sociotécnicos e plataformas digitais (pp. 59–72). Sociedade Brasileira de Estudos Interdisciplinares da Comunicação - Intercom.

Arriagada, A., & Ibáñez, F. (2020). “You Need At Least One Picture Daily, if Not, You’re Dead”: Content Creators and Platform Evolution in the Social Media Ecology. Social Media and Society, 6(3). https://doi.org/10.1177/2056305120944624

Bitencourt, E. (2023). Semantic Imagery Mapping (SIM): A method for analyzing the semantics of large image collections through computer vision [Computer software]. Datalab Design, Universidade do Estado da Bahia. https://datalabdesign.org/en/semantic-imagery-mapping

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. New Media & Society, 1164–1180. https://doi.org/10.1177/1461444812440159

Bucher, T. (2016). Neither Black Nor Box: Ways of Knowing Algorithms. In S. Kubitschko & A. Kaun (Eds.), Innovative Methods in Media and Communication Research (pp. 81–98). Springer International Publishing. https://doi.org/10.1007/978-3-319-40700-5_5

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. Information Communication and Society, 20(1), 30–44.

https://doi.org/10.1080/1369118X.2016.1154086

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), Advances in Knowledge Discovery and Data Mining (pp. 160–172). Springer. https://doi.org/10.1007/978-3-642-37456-2_14

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Deck, A., & Marasciulo, M. (2022, March 9). TikTok's biggest Chinese competitor bets big on Brazil. Rest of World. https://restofworld.org/2022/tiktok-competitor-kwai-brazil/

Flach, N. (2024, April 16). TikTok rival prepares expansion in Brazil. valorinternational. https://valorinternational.globo.com/business/news/2024/04/16/tiktok-rival-prepares-expansion-in-brazil.ghtml

Geboers, M. A., & Van De Wiele, C. T. (2020). Regimes of visibility and the affective affordances of Twitter. International Journal of Cultural Studies, 23(5), 745–765. https://doi.org/10.1177/1367877920923676

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417–441. https://doi.org/10.1037/h0071325

Jurno, A. C., & D'Andréa, C. F. de B. (2017). (In)Visibilidade algorítmica no "feed de notícias" do Facebook. Contemporânea | Comunicação e Cultura, 463–484.

Jokubauskaitė, E., Rieder, B., & Burkhardt, S. (2023). Winner-Take-All? Visibility, Availability, and Heterogeneity on Webcam Sex Platforms. Social Media + Society, 9(4), 20563051231214807. https://doi.org/10.1177/20563051231214807

Karhawi, I. (2024). Entre algoritmos, métricas de engajamento e plataformas digitais: Influenciadores digitais e trabalho de visibilidade. Revista Latinoamericana de Ciencias de la Comunicación, 23(46). https://doi.org/10.55738/alaic.v23i46.1144

Lin, J. (1991). Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory, 37(1), 145–151. https://doi.org/10.1109/18.61115

Lin, J., & De Kloet, J. (2019). Platformization of the Unlikely Creative Class: Kuaishou and Chinese Digital Cultural Production. Social Media + Society, 5(4), 2056305119883430. https://doi.org/10.1177/2056305119883430

Liu, K. Z. (2020). From invisible to visible: Kwai and the hierarchical cultural order of China's cyberspace. Global Media and China, 5(1), 69–85. https://doi.org/10.1177/2059436419871194

Luo, X., Cao, J., Sun, T., Yu, J., Huang, R., Yuan, W., Lin, H., Zheng, Y., Wang, S., Hu, Q., Qiu, C., Zhang, J., Zhang, X., Yan, Z., Zhang, J., Zhang, S., Wen, M., Liu, Z., Gai, K., & Zhou, G. (2024). QARM: Quantitative Alignment Multi-Modal Recommendation at Kuaishou (No. arXiv:2411.11739). arXiv. https://doi.org/10.48550/arXiv.2411.11739

Matamoros-Fernández, A., Gray, J. E., Bartolo, L., Burgess, J., & Suzor, N. (2021). What's "Up Next"? Investigating Algorithmic Recommendations on YouTube Across Issues and Over Time. Media and Communication, 9(4), 234–249. https://doi.org/10.17645/mac.v9i4.4184

McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction (arXiv:1802.03426). arXiv. https://doi.org/10.48550/arXiv.1802.03426

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50(302), 157–175. https://doi.org/10.1080/14786440009463897

Rieder, B., & Hofmann, J. (2020). Towards platform observability. Internet Policy Review, 9(4), 1–28. https://doi.org/10.14763/2020.4.1535

Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algorithms to 'ranking cultures' Investigating the modulation of visibility in YouTube search results. Convergence, 24(1), 50–68.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Strecker, H. (2024). Imaginários algorítmicos no Spotify: Sistemas de recomendação e a promessa de nos conhecer melhor do que nós mesmos. In Imaginários sociotécnicos e plataformas digitais (pp. 90–108). Sociedade Brasileira de Estudos Interdisciplinares da Comunicação - Intercom.

Tan, C. K., Wang, J., Wangzhu, S., Xu, J., & Zhu, C. (2020). The real digital housewives of China's Kuaishou video-sharing and live-streaming app. Media, Culture & Society, 42(7–8), 1243–1259. https://doi.org/10.1177/0163443719899802

Xi, R. (2024). Rural Chinese youth on Kuaishou: Performing gender, labor, and rurality. Journal of Youth Studies, 1–23. https://doi.org/10.1080/13676261.2024.2391934

Zhou, M., & Liu, S.-D. (2021). Becoming precarious playbour: Chinese migrant youth on the Kuaishou video-sharing platform. The Economic and Labour Relations Review, 32(3), 322–340. https://doi.org/10.1177/10353046211037090