Quantifying the Illicit Ecosystem of Betting Apps in India

Aatman Vaidya[1], **Kiran Garimella**[2]
[1]Tattle Civic Tech
[2]Rutgers University
aatmanvaidya@gmail.com, kiran.garimella@rutgers.edu

Abstract

Online betting and gambling apps in India have expanded rapidly, alongside growing concern about financial loss, debt stress, and addictive use patterns. Yet the ecosystem is difficult to quantify because recruitment and harm are observed in different places: users are often acquired through social media promotion, while harms become visible later inside apps and in user complaints. We address this measurement gap with a mixed-method, multi-source study that links promotion to downstream experience.

We compile three complementary datasets. First, we collect and analyze tens of thousands of betting-related advertisements from Meta's Ad Library using an extensive keyword strategy to measure scale and characterize persuasive frames. Second, we gather a purposive sample of organic Instagram posts from ten betting-linked hashtags to study how similar narratives circulate outside formal advertising, including through surrogate sports pages and influencer-style content. Third, we analyze over 300,000 Google Play reviews for a set of betting apps, using topic modeling to extract recurring user-reported problems that reflect a harm surface including financial loss, withdrawal friction, and customer support failure. We connect these layers by constructing a shared narrative codebook for paid and organic promotion and mapping those recruitment narratives to review topics.

Across sources, we find a consistent mismatch between what is promised at recruitment and what users report after adoption. Paid ads frequently frame betting as simple, quick, and highly winnable, while reviews repeatedly describe difficulty winning, blocked or delayed withdrawals, unclear rules, and perceived extractive design. Organic promotion often uses more coded and informal presentation than official ads, potentially reducing detectability while funneling users toward the same apps and referral pathways. Together, these results provide one of the first large-scale, cross-source measurements of India's online betting promotion ecosystem and its associated user-reported harms, and they offer a general approach for studying how potentially harmful digital services sustain growth through mainstream platforms even under evolving regulation.

1 Introduction

Online betting and gambling apps in India have expanded quickly, while the social costs have become harder to ignore.

News reporting and regulatory debate increasingly connect these products to severe financial loss, debt stress, and addictive use patterns, especially among young and lower-income users who face high exposure to sports-linked marketing and easy in-app payments (Bhatt 2022; Tiwari 2023; Krishnan 2025). At the same time, the market has been shaped by a long-standing legal and enforcement mismatch. Gambling is regulated primarily at the state level, and many platforms position themselves as "games of skill" or as entertainment products rather than gambling, while offshore operators can target Indian users from outside jurisdiction and re-enter after takedowns through new domains, brand shells, or surrogate pages (India Today 2025; Business Standard 2025).

This paper argues that the key empirical challenge is not only what these apps do, but how they recruit. For most users, the entry point is not an app store search, but a deluge of promotion on mainstream social platforms: short-form ads, influencer endorsements, and "lifestyle" posts that frame betting as a low-risk extension of sports fandom or a viable path to quick income. Although India's 2025 legislative efforts aim to curb real-money gaming ads (Singh 2025), the ecosystem is highly adaptive. When explicit ads are blocked, promotion shifts to "surrogate" content—disguised as sports news, memes, or fan pages—bypassing platform policies that technically prohibit fraudulent advertising but functionally fail to stop it (Financial Express 2025; Horwitz 2025; Gandhi 2026).

Measuring this ecosystem is difficult because standard approaches often miss the mechanism connecting recruitment to harm. Studies that focus solely on app forensics miss the persuasive messaging that drove the download; studies that focus solely on ad archives see the sales pitch but not the downstream financial ruin. Furthermore, organic content spreads through informal, ephemeral networks that are difficult to scrape at scale, while user reviews, though plentiful, are often noisy and skewed. Consequently, prior work has remained largely single-sited, analyzing either the legal definitions, isolated marketing campaigns, or clinical health outcomes, without connecting the dots between them (Benegal 2013; Hoonka 2022; Krishnamoorthy et al. 2025).

We address this fragmentation with a mixed-method, multi-source design that treats *promotion* and *harm* as linked and measurable stages. First, we analyze tens of thousands of betting-related advertisements collected from Meta's Ad

Library using an extensive keyword strategy. This corpus allows us to quantify scale and to characterize dominant persuasive frames at scale. Second, we study a purposive sample of organic Instagram content by collecting posts from ten betting-linked hashtags, capturing how similar narratives appear when promotion is embedded in everyday posting rather than formal ad placement. Third, we analyze over 300,000 Google Play reviews for a set of betting apps, using topic modeling to extract recurring user-reported problems. We interpret these topics as a harm surface that includes financial loss, withdrawal friction, customer support failure, and signals consistent with problematic use. We connect these layers by building a shared narrative codebook across paid and organic social content and then mapping those coded narratives to review topics, allowing us to test where recruitment promises and user complaints align, and where they diverge.

Our results reveal a consistent mismatch between recruitment claims and reported experience. In the Meta ad corpus, a dominant set of narratives presents betting as simple, quick, and highly winnable, often emphasizing low effort and high returns. In contrast, Google Play reviews repeatedly describe difficulty winning, blocked or delayed withdrawals, unclear or shifting rules, and a sense that the system is designed to extract deposits rather than enable payouts. Organic Instagram content, packages promotion differently from official ads, relying more on informal cues, coded language, community slang, and influencer-style presentation that may reduce the risk of automated detection while still driving users toward the same apps and referral pathways. Though these findings do not establish causal effects on individual users, triangulation across paid promotion, organic dissemination, and user reports provides evidence of a coordinated acquisition ecosystem and a large, measurable harm footprint in the post-adoption experience.

2 Related Work

Gambling-related harm and the Indian context. Public health research increasingly frames online gambling as a population-level risk shaped by product design, frictionless payments, and pervasive digital promotion (Hing et al. 2022; Tiwari 2023). The Lancet Public Health Commission emphasizes that harms extend beyond clinically defined disorder and can affect families and communities through financial strain, relationship breakdown, and mental health consequences, with digital access and marketing expanding both reach and intensity (Wardle et al. 2024). Evidence reviews also argue that advertising exposure is plausibly linked to gambling participation, with heightened concern for children, young people, and those already at risk (McGrane et al. 2023; Hing et al. 2014). In India, scholarship highlights limited epidemiological measurement and the regulatory mismatch created by state-level governance and the skill-versus-chance distinction (Benegal 2013). Recent studies of fantasy sports in India similarly point to widespread use and contested perceptions of whether products are gaming or gambling, reinforcing the need to measure real-world recruitment and harm rather than relying only on legal categories or self-report (Balhara, Gulati, and Rajguru 2024).

Gambling promotion on social media. A second literature examines how gambling promotion operates on social platforms. Content analyses show that marketing is often blended with sports and entertainment content, humor, and interactive engagement, which can reduce the salience of persuasive intent and complicate enforcement (Stadder and Naraine 2020). Work also emphasizes the role of third-party affiliates and creators: affiliates can be more direct and transactional than official brand accounts, while operators may rely more on branding and engagement, creating a division of labor that pushes the most explicit recruitment pressure outside official channels (Houghton et al. 2019). Reviews argue that measurement needs to account for this heterogeneity across actors and platforms (Singer and Wöhr 2024). Related work on influencer marketing and disclosure is relevant because disclosure norms shape whether audiences interpret content as community talk or paid persuasion (Bertaglia et al. 2025), and emerging evidence documents persistent influencer-driven gambling exposure among young people even under nominal disclosure and age-gating rules (Pitt et al. 2024). These findings motivate our design choice to analyze both paid ads and organic social content, since they can carry similar narratives but in different wrappers.

Auditing ad ecosystems and platform governance. A broader accountability literature uses advertising transparency tools (such as Meta and Google ad libraries) as partial observational windows into advertiser behavior and platform enforcement. Audits of political advertising and other restricted domains show that policy statements and transparency mechanisms do not guarantee effective enforcement in practice, and that systematic measurement can reveal persistent gaps (Edelson, Lauinger, and McCoy 2020; Le Pochat et al. 2022). Other work highlights structural limitations of ad libraries, including incomplete coverage and limited visibility into targeting and exposure, and proposes complementary approaches such as user-donated data for stronger inferences (Gkiouzepi et al. 2023; Andreou et al. 2019). Recent studies similarly use transparency repositories to quantify potentially illegal or non-compliant advertising while noting that these data alone cannot establish exposure or downstream harm (Xiao 2025). This motivates multi-source designs that pair ad-library measurement with other lenses.

Evasion and organic adaptation. Even when paid advertising is constrained, promotion can persist through organic channels that are harder to police. Research on content moderation documents how communities evade detection through code words, creative spelling, and shifts in style and framing (Gerrard 2018). While these studies are often grounded in other domains, the mechanism generalizes: when enforcement relies on automated detection and simple triggers, actors can adapt presentation without changing underlying intent (Casado-Aranda, Dimoka, and Sánchez-Fernández 2019). In the gambling context, this suggests that organic promotion may diverge from paid ads not because the product differs, but because the content is optimized to survive moderation and spread socially.

3 Data Collection

This study is designed as a single, end-to-end measurement of the betting app ecosystem as it is encountered by users in practice. The core idea is to observe three connected surfaces of the same phenomenon: (i) paid acquisition at scale (advertisements), (ii) organic or semi-organic social diffusion (content on Instagram), and (iii) downstream user-reported consequences after adoption (app store reviews). Rather than treating these as independent datasets, the collection was structured around a common set of entities and narratives. We began by qualitatively mapping the language and visual tropes used to promote betting, used that mapping to collect paid advertisements at scale, and then followed the same promotional pathways to identify the apps being pushed and the user complaints associated with them. This triangulated design aims to capture a realistic lifecycle of exposure, normalization, and reported harm, while remaining explicit about the observational limits of each data source.

We first conducted an exploratory qualitative review to identify the vocabulary, imagery, and referral pathways commonly used to promote betting and gambling content. Using a snowball sampling approach, we started with a small set of seed keywords and conducted hashtag searches on Instagram. We examined the resulting posts, the accounts they linked to, and the off-platform destinations they promoted, including Telegram groups and third-party websites (Cruickshank and Kloo 2023; Atkinson and Flint 2001). As new terms, app names, and coded phrases surfaced, we iteratively expanded the keyword list to improve coverage. The final keyword list is reported in Appendix A.1.

3.1 Meta advertisements

To measure the scale and characteristics of paid promotion, we collected advertisements using the Meta Ad Library API[1]. Data were collected mid 2025 using the finalized keyword list. Because the API supports querying using a minimum date constraint, we collected ads in temporal batches to ensure coverage across multiple years. Specifically, we issued successive queries with progressively earlier minimum dates, retrieving ads spanning 2023 through 2025. This procedure yielded approximately 340,000 unique ads prior to filtering.[2]

Initial inspection revealed substantial keyword spillover and unrelated content. A common failure mode was substring matching, where keywords containing strings such as "win" retrieved unrelated promotions (for example, fictional story or casual gaming apps). To address this, we conducted systematic filtering based on ad text, landing-page cues when available, and manual inspection of high-frequency unrelated content clusters. We retained ads that plausibly promoted betting, gambling, or real-money gaming services, resulting in 29,772 ads. Table 3 (Appendix) summarizes the removed unrelated content categories.

For ads run in India, the Ad Library provides rich metadata only for certain categories such as political and social issues, and most ads in this corpus did not include reliable reach or spend statistics.[3] For this reason, analysis focuses on the media and messaging of the ads as a proxy for potential harm and persuasion, rather than on who was targeted or the spend amount. Each ad record contains a snapshot URL to an HTML page. We used Selenium-based tooling to retrieve the creative content from these pages and download associated images and videos for offline analysis. Media were successfully downloaded for 23,620 ads, comprising 14,069 images and 9,551 videos. The remaining ads could not be downloaded due to missing or dynamically loaded elements and access instability at the time of collection. Most of the ads were short-lived, with over 60% of the ads active for only 1 to 3 days, consistent with rapid campaign turnover and possible attempts to reduce enforcement exposure.

3.2 Instagram posts

To study how betting promotion manifests outside the 'top down' content push from the betting app companies, we collected a purposive sample of public Instagram posts using hashtag search. We selected 10 hashtags from the finalized keyword set, prioritizing terms that appeared frequently during the exploratory review and that had high visible posting volume. For each hashtag, we collected up to 200 posts, yielding 2,695 posts in total. Collection was performed using Selenium-based scraping tools that interacted with public web interfaces and stored only what was necessary for content analysis.

Given platform constraints and the absence of a stable public API for large-scale research collection of Instagram content, this dataset is intentionally designed for qualitative and comparative analysis rather than prevalence estimation. We downloaded only image content and stored the post identifier needed for deduplication. In particular, the Instagram sample is used to compare how promotion is packaged in organic contexts relative to paid ads, not to measure the overall rate of exposure on Instagram. Instagram posts frequently contained references to specific app names, referral links, and off-platform coordination channels. The next subsection describes how these links were operationalized to define the set of betting apps for which downstream user reviews were collected.

To gain a deeper understanding of the broader Instagram ecosystem, we conducted a qualitative review of both posts and their associated user profiles. Across all 10 hashtags, we manually examined around 200 Instagram posts and profiles. For each entry, we recorded the post ID, username, follower count, number of posts, engagement rate, and manual notes describing the user's overall content, as well as the potential harms associated with it. These observations were documented in a structured Google Sheet for analysis.[4]

[1] https://www.facebook.com/ads/library/api/

[2] Meta's Ad Library terms permit research use of the API for studying advertising and publishing reports or academic work.

[3] This restriction surprisingly does not apply for ads run in the European market.

[4] Link to the Google Sheet with our collected and analyzed data for the Qualitative Review: https://bit.ly/icwsm-betting-apps

3.3 Google Play Store reviews (downstream user-reported consequences)

To capture user-reported experiences after adoption, we collected Google Play Store reviews for betting-related apps promoted in the two social datasets collected above. During review of the Meta ads and Instagram posts, many creatives pointed directly to Google Play listings or referenced app names that could be matched to Play Store packages. Among the apps linked in advertisements and posts, we identified 58 apps relevant to real-money betting or casino-style play and collected their one-star reviews as a focused lens on negative experiences. Reviews were collected using an open-source scraper[5] and resulted in 329,501 one-star reviews.

This review dataset is not treated as a representative sample of all users. Instead, it is used as a large-scale repository of harm narratives and product-friction reports that are difficult to observe from the outside. Restricting to one-star reviews increases the density of complaints related to loss, withdrawal failures, account restrictions, customer support breakdowns, and perceived deception, which are central to the study's harm framing. At the same time, this choice introduces well-known selection effects: satisfied users are underrepresented, some reviews may be impulsive or strategic, and the most severe harms may occur among users who never leave a review. These limitations are addressed in analysis by focusing on robust, recurring themes rather than treating any single complaint as ground truth.

3.4 Summary and ethical considerations

Overall, the final dataset consists of 23,620 Meta ad creatives (from 29,772 retained ad records), 2,695 Instagram posts, and 329,501 Google Play one-star reviews[6]. The three sources are linked by construction through a shared discovery process and shared entities: keywords and coded terms surfaced during exploratory review informed ad collection; ad creatives and organic posts revealed the apps and referral pathways being promoted; and those same apps defined the Play Store review corpus.

Throughout collection, the goal was to measure ecosystem-level patterns while minimizing privacy risk. Data collection focused on publicly accessible content and platform-provided transparency interfaces, stored only the fields needed for analysis, and avoided collecting personal identifiers or engagement traces on Instagram. For the app store, reviews are public user submissions, but analysis is conducted at the aggregate level to characterize themes rather than individuals.

4 Methods

4.1 Annotation of Meta Ads and Instagram Posts

To characterize the nature and potential harms of the advertisements in our dataset, we manually annotated all 23.6k Meta ads. We adopted a bottom-up, inductive coding approach (Glaser and Strauss 2017; Hwang, Nanayakkara, and Shvartzshnaider 2025). One author initially started by annotating a subset of ads, taking into account the ad media content, description, and any external websites linked from the ad. A special annotation interface was designed to help with the annotation task (see Figure 2 for the annotation UI). This helped identify recurring themes and sources of potential harm. We then iteratively updated the annotation guidelines, adding new betting categories and additional harm types as they emerged during the entire annotation process.

The final annotation guideline comprised of the following questions:

1. **Unrelated Ads Identification:** We first indicated whether the ad appeared to be unrelated to betting apps. This distinction allows us to separate genuine promotional content from unrelated advertisements that do not meaningfully pertain to gambling or betting.
2. **Ad Category:** Ads were categorized into domains such as sports betting, fantasy sports, casino games, card games, prediction games, lotteries/jackpots, or trading/crypto. This captures the type of gambling-related activity being promoted.
3. **App Name:** The name of app/website being promoted (e.g., 1xBet, Betway, Parimatch, Bet365), with an other option when not listed. This allows us to link ads to particular platforms.
4. **Primary Messaging Strategy:** We annotated one or more dominant persuasive strategies used in the ad, like user acquisition, promotion of a new feature, celebrity endorsement, misleading or deceptive claims, emotional appeal, social proof/testimonials, or general promotion. This helps us understand how betting platforms market themselves and what rhetorical strategies are most commonly employed.
5. **Potentially Harmful Narratives:** We captured the presence of harmful themes known to increase risk or vulnerability, such as easy money, risk minimization, income-substitute framing, normalization of addictive behavior, debt-solution framing, manipulated success stories, or time-pressure tactics.
6. **Media Authenticity:** We identified whether the visual content appeared authentic or AI-generated. Given the increasing use of generative models in advertising (Baek 2023), this helps quantify the prevalence of synthetic media in gambling promotion. We define AI-generated content as an image, audio, or video content that has been automatically synthesized or manipulated to depict people saying or doing things they never did (Farid 2025).
7. **Sexual Content:** Does the ad contain sexualized content or themes, like soft porn or erotic content.
8. **Ad Notes:** Text field to record additional observations.

A detailed version of the annotation guideline and options for each question are provided in Appendix A.2.

One of the authors manually annotated all the 23.6k advertisements. Early in the process, we observed that a substantial number of ads shared visually similar media con-

[5]https://github.com/JoMingyu/google-play-scraper

[6]The full dataset is publicly available at https://github.com/aatmanvaidya/Quantifying-the-Illicit-Ecosystem-of-Betting-Apps-in-India.

tent. To reduce redundancy in the annotation workload, we applied PDQ and TMK+PDQF perceptual hashing[7] (Farid 2021) to identify near-duplicate images and videos. Using a 95% similarity threshold, we clustered visually similar media items and reduced the number of items requiring manual annotation from 23.6k ads to 9,892 unique media items. We then annotated these items using the guidelines described in Section 4.1. The assigned labels were propagated to all ads belonging to the same similarity cluster. We report the aggregate annotation results for the Meta Ads in Table 1.

Overall, 55.41% of the ads were labeled as unrelated. Despite the earlier automatic filtering described in Section 3, unrelated advertisements remained present in the dataset. These primarily consisted of non–betting related content such as generic e-commerce promotions, fictional story-app advertisements, political campaign material, and unrelated service advertisements. In Table 1, we report the annotation results for remaining 44.59% i.e. 10,526 relevant ads.

In order to scale the annotation to further datasets, using the existing manual annotated corpus of ads, we applied a few-shot classification approach using `gemini-2.5-pro` to the Instagram dataset. The model was provided with our annotation guidelines (see Appendix Section A.3 for the detailed prompt and several labeled few-shot examples). We tested the model's performance using a held out set, we randomly sampled 300 images and the model performed well achieving 97% accuracy. Annotations were considered correct when the model accurately identified the ad category, potentially harmful narratives, media authenticity, and broadly captured messaging strategies.

4.2 Topic Modeling on Play Store Reviews

To characterize downstream user-reported harms at scale, we performed topic modeling on all one-star Google Play reviews using BERTopic (Grootendorst 2022). We focus on one-star reviews because they concentrate complaint narratives, allowing recurring failure modes such as withdrawal problems, perceived deception, loss experiences, and service breakdowns to surface with higher density than in mixed-rating corpora. Topic modeling is used here as a descriptive summarization tool. It does not establish causal effects or measure prevalence of harms in the full user population, but it provides a structured view of the dominant complaint themes expressed by users who chose to leave negative feedback.

We first normalized review text by lowercasing, converting emojis to their textual descriptions (de-emojizing), and removing URLs. We retained the remaining text largely as written, avoiding aggressive linguistic normalization such as stemming or lemmatization because BERTopic clusters reviews primarily through semantic embeddings rather than exact word matching. The cleaned texts were embedded using the `all-MiniLM-L12-v2` sentence-transformer model, chosen for its strong performance on short-text semantic similarity and its computational efficiency at the scale of our corpus.[8]

[7]https://github.com/facebook/ThreatExchange

[8]Topic modeling was run on a Tesla P100 16GB GPU provided by Kaggle Notebooks.

We then configured BERTopic with a standard pipeline that combines embedding-based clustering with interpretable token-based representations. We set the vectorizer to `min_df=2` and `ngram_range=(1,2)` with English stop words, used UMAP for dimensionality reduction with cosine distance (`n_neighbors=15`, `n_components=5`, `min_dist=0.0`), and applied HDBSCAN for clustering. Topic descriptors were derived using BERTopic's c-TF-IDF representation, which provides salient n-grams for each cluster and helps interpret themes even when clustering is driven by dense embeddings.

To make topics usable in the narrative analysis, we generated human-readable topic labels with the `Qwen2.5-7B-Instruct-1M` LLM by prompting it with each topic's top n-grams and representative review snippets. We validated these labels by randomly sampling 50 reviews from each of 15 topics spanning both high- and mid-frequency clusters and manually checking whether the label accurately described the dominant content. When a label was too broad or inconsistent with the samples, we revised it to reflect the most common complaint expressed in that cluster.

5 Results

5.1 Social promotion across paid and organic channels

Table 1 reports a unified view of betting-related content on two complementary social surfaces: paid acquisition on Meta and organic (or semi-organic) diffusion on Instagram. Reading these results together is important because the platforms occupy different positions in the distribution chain. Meta ads reflect marketing that must pass through a formal ad review process and is optimized for conversion. Instagram hashtag posts reflect content that circulates through social discovery and influencer norms, where promotion can be embedded in entertainment, advice, or community talk. The quantitative patterns show both a shared persuasive core across platforms and systematic differences in how that core is expressed.

At the level of product categories, the clearest cross-platform signal is what can be described as the *financialization* of paid ads. Meta's ad corpus contains a substantial share of content framed as finance-adjacent rather than explicitly "gambling": trading/crypto accounts for 20.05% of relevant ads and lottery/jackpot formats account for 18.90% (Table 1). By contrast, trading/crypto is nearly absent in the Instagram hashtag sample (2.42%). This asymmetry suggests that paid advertising relies heavily on camouflaged gambling, where products borrow the language and aesthetics of legitimate finance, automated "trading," or skill-based earning to expand appeal beyond self-identified gamblers and to reduce the chance that the creative is immediately recognized as prohibited betting promotion. In comparison, the organic Instagram sample is more direct and less camouflaged: it is dominated by overt casino-game content (71.93%) and substantial sports betting (32.23%), alongside prediction games (24.28%) (Table 1). This pattern is consistent with the idea that hashtag-based discovery disproportionately captures audiences already searching for, following, or embedded in gambling-related communities, where explicit gambling vernacular is both legible and socially reinforced, and therefore provides less incentive to disguise the underlying activity as "finance" or "investment."

The messaging strategies further highlight this split. The Meta corpus is industrialized for immediate conversion: nearly all ads (99.7%) focus on "User Acquisition," and 85.9% emphasize specific features (e.g., "instant withdrawal," "sign-up bonus"). This "feature-first" approach is designed to reduce friction and trigger a deposit. Conversely, the Instagram dataset relies more on "General Promotion" (95.6%) and utilizes Celebrity Endorsement (9.9%) and Emotional Appeals (9.6%) far more frequently than in ads (<2%). [9]

The potentially harmful narratives reinforce this picture of a shared persuasive core with different expression levels across channels. On both platforms, two frames are nearly universal: easy money and risk minimization, indicating that the promise of earnings and the downplaying of risk travel across both surfaces, even when the content is not formally an advertisement (Table 1). The channels diverge in the specific harm variants that are emphasized. Meta ads more often frame betting as an income stream (23.40% vs. 8.59%), consistent with a conversion-oriented pitch that positions participation as livelihood rather than leisure. Instagram, by contrast, shows higher rates of success-story manipulation (11.43% vs. 1.85%), addiction normalization (9.74% vs. 0.33%), and time-pressure tactics (3.49% vs. 0.05%). One interpretation for this stark difference is that content that looks overtly coercive or socially sensitive is less compatible with paid ads and therefore shifts to organic channels, where such cues can be expressed through community norms and informal language (Tiwari 2023). This interpretation is consistent with the broader pattern that betting-related promotion persists despite Meta's stated restrictions on gambling advertising and its general prohibitions on deceptive practices.[10]

Finally, we observe distinct patterns in how credibility is manufactured across the two channels. On Meta, where organic social context is absent, advertisers appear to compensate by synthetically engineering trust: a substantial 14.84% of paid creatives contained AI-manipulated media, compared to only 4.38% in the organic Instagram sample. This reliance on synthetic authority correlates with a broader strategy of deception in the paid tier, where 67.38% of ads were flagged as misleading or deceptive. We use the term "AI generated" to refer to manipulated audio and/or video that makes a real person appear to say or do something they did not, synthesized primarily by generative models rather than simply edited from real-world recordings. In practice, these manipulated creatives are deployed to simulate authority and fabricate endorsements; several ads feature "news-like" overlays or deepfakes of public figures to imply legitimacy (e.g., Figure 1(g)). We observed deepfakes depicting individuals promoting apps that allegedly allow users to earn money daily, as well as repurposed clips of well-known personalities with new overlaid text associating them with betting platforms. Notably, several deepfakes featured Indian celebrities promoting the crash prediction game *Aviator*, including internet personality CarryMinati, cricketer Virat Kohli (see Figure 1(f)), and business figures such as Mukesh Ambani and Anant Ambani, all shown encouraging sign-ups and claiming that substantial financial gains are easy. The use of celebrity deepfakes for *Aviator* has been documented by fact-checkers and media forensics in India,[11] and journalistic reporting further connects *Aviator* use to significant user losses.[12]

Overall, the quantitative results support a unified account of an ecosystem with a stable persuasive core (earnings plus safety) that adapts to platform constraints: explicit feature selling, deception, and synthetic credibility cues are more visible in paid advertising, while organic content relies more heavily on broad promotional framing and socially mediated persuasion.

5.2 Qualitative Analysis

The quantitative patterns in Table 1 become more legible when paired with close reading of representative content and distribution pathways. Across both Meta ads and Instagram posts, the ecosystem repeatedly relies on the same persuasive template: foreground a vivid signal of winning, minimize any perceptible risk, and then route users toward an off-platform conversion point (Telegram groups, third-party sites, or an app install). The qualitative review in this section helps explain how this template is operationalized, and why it can remain effective even when platforms state restrictions on gambling promotion and deceptive conduct.[13]

In the Meta ads, many casino-style creatives do not clearly disclose an app, operator, or legal entity; instead, they foreground slot-machine wins, large balances, and joining bonuses (e.g., "200% bonus"), which functions as an implicit guarantee that earnings are both likely and repeatable. Sports betting ads similarly center on screenshots of winning slips and "high-confidence" predictions (Figure 1(a,c,h)),

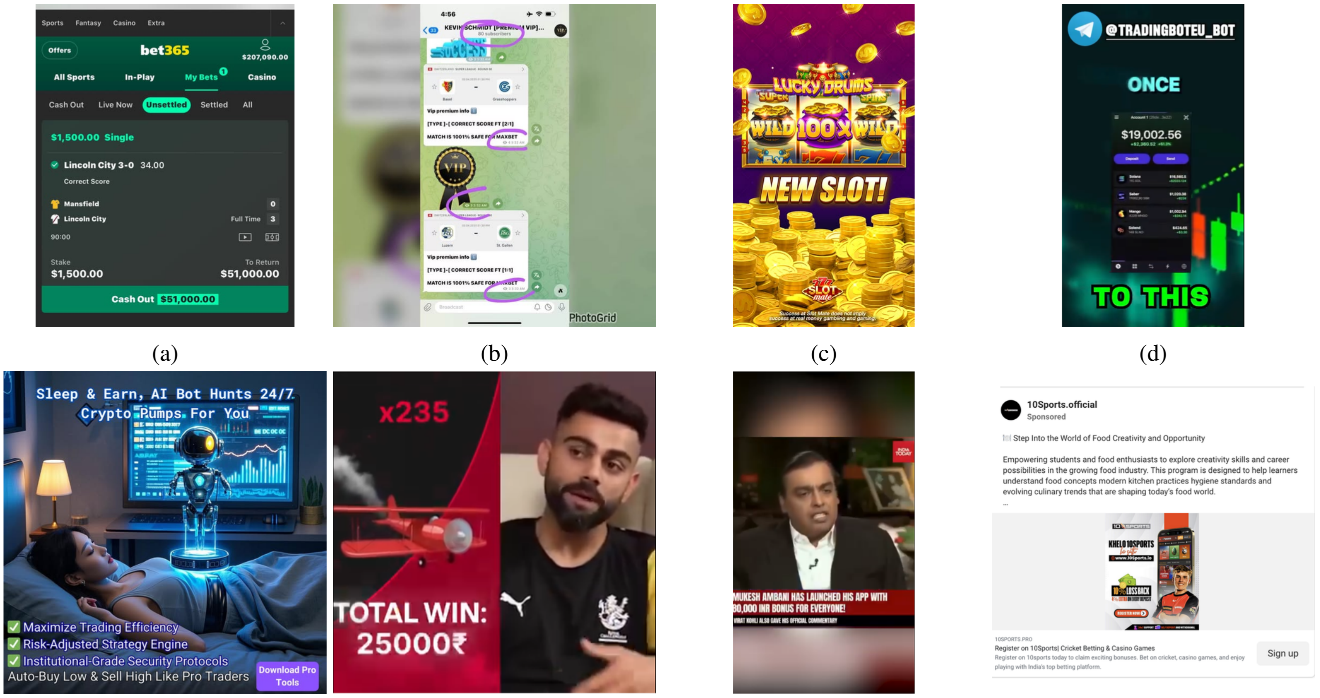
[9] We note a methodological caveat: The discrepancy between "Feature Promotion" (high on Meta) and "General Promotion" (high on Instagram) may be partially attributable to the difference between manual annotation (Meta) and LLM classification (Instagram), as the model tended to classify gameplay mechanics as general promotion. However, the qualitative difference in tone of transactional vs. lifestyle remains consistent upon manual review.

[10] Figure 3 reproduces the relevant Meta policy language; see https://transparency.meta.com/policies/ad-standards/#restricted-goods-and-services and https://transparency.meta.com/policies/community-standards/fraud-and-scams/.

[11] https://www.boomlive.in/fact-check/virat-kohli-aaj-tak-shweta-singh-betting-app-aviator-deepfake-scam-26523, https://www.boomlive.in/decode/fraud-ads-ai-voice-clones-deep-fake-facebook-india-apps-laila-rao-shahrukh-khan-ravish-kumar-deep-fake-akshay-kumar-23815

[12] https://www.boomlive.in/decode/aviator-youtubers-ai-meta-boost-an-illegal-game-india-28143

[13] See Figure 3 (Appendix) for a screenshot of Meta's explicit policy against advertising and posting such content on their platforms.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 1: Examples of the Meta Ads and Instagram posts. (a) Sports Betting ad promoting the Bet365 app, showcasing a large winning payout on a bet. (b) Sports Betting ad promoting a Telegram channel providing betting tips, shows high-confidence predictions that minimize the perception of risk and may be misleading to users. (c) Casino Game ad aimed at user acquisition and general promotion, illustrating gameplay and large winnings to suggest that earning money is easy. (d) Crypto/Trading ad featuring a trading bot, showing potential high earnings and promoting automated trading. (e) AI-generated Crypto/Trading ad promoting an automated trading bot, shows high potential earnings while downplaying associated risks. (f) Deepfake video ad featuring the sports personality Virat Kohli promoting the Aviator prediction game, shows potential earnings from chance-based games, which may create unrealistic expectations of success. (g) Deepfake video ad showing businessman Mukesh Ambani with a fabricated news banner to promote an app offering an 80,000 INR bonus, conveying misleading endorsements. (h) A Sports Betting app ad with misleading text about cooking while the link points to a casino app.

spanning widely watched leagues as well as smaller regional competitions. This “proof-first” style is important because it shifts the viewer’s attention away from probabilistic loss toward a concrete image of success, creating a perception that outcomes are visible, knowable, and therefore controllable. In ads that promote prediction or crash-style games, the same structure is preserved through short, procedural demonstrations that make winning appear mechanical rather than uncertain, often implying that there is a method or hidden rule that reliably converts deposits into payouts.

We observed a recurring operational pattern in Meta off-platform conversion: Ads routinely instruct users to move to Telegram channels for betting tips, to install apps through Google Play, or to visit third-party websites (Figure 1(b,d)). We sampled a subset of Telegram groups linked from ads and found that many are organized around a specific betting form (e.g., football tips) or around a specific platform. These groups present tips as near-certain, with repeated language that minimizes risk and frames uncertainty as solvable through insider knowledge. The effect is to outsource the highest-pressure persuasion to spaces that are less visible to platform oversight and easier to regenerate after enforcement. The trading/crypto category follows a parallel logic: gambling-like activity is reframed as investment or income generation, often using luxury imagery and “automation” rhetoric (Figure 1(e)).

The qualitative review of Instagram posts complements this picture by showing how the same ecosystem is sustained through accounts and content that are structurally closer to “community media” than to formal advertising. For each hashtag corresponding to our keyword set, we randomly sampled 20 posts for in-depth inspection and then visited the posting profiles to analyze the surrounding account behavior. The profile-level pattern was similar for most accounts, betting- and gambling-related material constituted nearly all of their posts rather than appearing as an occasional topic. Many profiles appeared purpose-built for gambling promotion, with consistent posting styles, repeated motifs, and persistent calls to move to an external destination. To a large extent (around 90%), the accounts involved were not organic or undisclosed promotion but an organized set of dedicated promotion accounts that resemble a well resourced, top down distribution infrastructure. These accounts also do not have a high follower count or engagement with follower

Annotation Category	Meta Ads (%)	Instagram (%)
Ad Category		
Casino Games	49.45%	71.93%
Sports Betting	26.61%	32.23%
Trading / Crypto	20.05%	2.42%
Lottery / Jackpots	18.90%	9.13%
Prediction Games	17.43%	24.28%
Card Games	4.19%	3.99%
Other	1.25%	0.19%
Fantasy Sports	0.49%	5.76%
Primary Messaging Strategy		
User Acquisition	99.77%	95.06%
Promoting a New Feature	85.90%	0.96%
Misleading or Deceptive	67.38%	14.08%
Social Proof / Testimonials	59.09%	7.79%
General Promotion	3.50%	95.67%
Celebrity Endorsement	1.62%	9.94%
Emotional Appeal	0.56%	9.63%
Potentially Harmful Narratives		
Easy Money Narrative	99.39%	79.33%
Risk Minimization	98.40%	78.24%
Income Source Framing	23.40%	8.59%
Debt Solution Marketing	1.85%	0.42%
Success Story Manipulation	1.85%	11.43%
Other	0.79%	0.26%
Addiction Normalization	0.33%	9.74%
Time Pressure Tactics	0.05%	3.49%
Media Authenticity		
Authentic	85.21%	95.62%
AI-Generated	14.84%	4.38%
Sexual Content		
No	99%	98.8%
Yes	1%	1.2%

Table 1: We report annotation results for betting and gambling-related content on Meta Ads and Instagram. Full annotations were not performed on unrelated content. One author manually annotated all 10,526 relevant Meta advertisements, while the 2,607 related Instagram posts were annotated using few-shot classification with Gemini-2.5-Pro. We report the results in percentages, these are calculated with respect to the total number of relevant annotated items

counts in the range of 1K–5K, suggesting limited genuine audience interaction.

The qualitative review also surfaced channel-specific credibility tactics. Instagram profiles frequently relied on social identity and repeated posting rather than explicit feature claims, but many still used manufactured authority cues. We observed celebrity imagery and endorsement cues, both real and synthetic, including apparent deepfakes or repurposed clips of sports personalities used to imply legitimacy. Some accounts leaned into sexualized or erotic imagery and, in a few cases, used adult entertainers as attention anchors for promotion. These tactics were qualitatively more visible in Instagram than in Meta ads, where promotion more often remains within a narrow persuasive frame of winnings, bonuses, and "trusted" status.

At the level of content, Instagram posts reproduce many of the same themes visible in Meta ads, including easy money narratives, risk minimization, promised high returns, and depictions of frequent or effortless wins. The difference is less in the underlying claim and more in how that claim is wrapped. Compared to the Meta ads, Instagram posts more often embed the pitch in memetic forms, short clips, and informal presentation that reads as entertainment or "tips" rather than as an explicit product offer.

5.3 Google Play Reviews: The Harm Surface

The social promotion analyzed in the previous sections is only one part of the ecosystem's impact. To understand what happens after adoption, we analyze one-star Google Play reviews for the set of betting-related apps identified through the social datasets. Reviews are not a representative survey of all users, but they provide a large-scale, user-proximate record of what people report when an app fails them, frustrates them, or harms them. Topic modeling helps summarize this record by surfacing recurring complaint clusters across hundreds of thousands of short texts. The BERTopic procedure produced 220 topics, which we consolidated into broader harm themes by grouping semantically similar clusters (e.g., multiple variants of "withdrawal pending," "unable to cash out," or "login failure"). Table 2 reports the distribution of reviews across these consolidated themes.

Broader Topic Theme	Review Count
Financial Loss and Scam	95,158 (28.88%)
Miscellaneous Negative Reactions	59,626 (18.10%)
Technical Problems and Reliability Issues	42,044 (12.76%)
User Experience and Usability	39,331 (11.94%)
Gameplay Fairness and Bots	38,846 (11.79%)
Miscellaneous Positive Reactions	15,226 (4.62%)
Advertising and App Permissions	13,805 (4.19%)
Customer Service Issues	12,630 (3.83%)
App Updates and Version Related Issues	9,591 (2.91%)
Social and Multiplayer Features	3,244 (0.98%)

Table 2: Distribution of user reviews by consolidated thematic topic. Note the dominance of financial complaints and negative sentiment.

A first headline result is that the most common theme is explicitly financial harm. The single largest category is Financial Loss and Scam (28.88%), comprising nearly 100,000 reviews. When combined with Miscellaneous Negative Reactions (18.10%) and Gameplay Fairness (11.79%), we find that nearly 60% of the reviews consists of explicit complaints regarding loss, unfairness, or deception.

Beyond the volume of complaints, qualitative analysis reveals *how* these harms manifest. Users describe a sophisticated system of "bait-and-switch" mechanics that directly contradict the “fair play” narrative seen in ads. A recurring narrative within the Financial Loss and Gameplay Fairness topics is the perception that game odds are not static but reactive to user spending. Users frequently report a pattern where they are allowed to win initially (the “hook”), but face impossible odds immediately after making a real-money purchase. This suggests users perceive a form of predatory *dynamic difficulty adjustment* designed to extract capital.

```
Don't spend any money. As soon as
you make a purchase you will always
lose.
```

```
It'll give you a ton and then take
it all so you have to buy... I
bought the 7 million chips package
for $11.99, and when I opened the
game again today, I was at 4,000
chips.
```

Unlike a standard casino where the house edge is statistical and constant, users describe these apps as actively adversarial, leading to a sense of betrayal:

```
This game reminds me why I never
gamble IRL... The 1% unfavorable
probabilities happen quite often in
this game, to push you to buy chip
packs.
```

While the Technical Problems topic (12.76%) might appear to be benign software bugs, qualitative inspection suggests many users perceive them as intentional friction points. Complaints often describe apps that function perfectly during deposits but “crash” or “freeze” during winning hands or withdrawal attempts.

```
It keeps disconnecting... I
couldn't see my cards so I got
angry and closed the app.
```

```
For some reason it doesn't respond
to my raise or check sometimes. I
would have won 2 hands in the 30k
and it made me fold.
```

Perhaps the most alarming signal in the dataset is the prevalence of severe psychological distress. While the ads frame betting as “leisure” or “skill,” the reviews document a transition into compulsion, anger, and clinically significant mental health strain.

Users explicitly recognize the addictive design of the interfaces, often describing a loss of control. Notably, these harms spill over to vulnerable family members, corroborating concerns about youth exposure.

```
It is addictive and my son played
for first time and spend over
20,000 on chips it is crazy.
```

```
Set up hands..so that we lose chips
so fast ..so that we will buy chips
for money.....addictive tho.
```

The emotional spectrum of the reviews shifts from anger to deep despair. We identified a cluster of reviews that explicitly link app usage to “depression,” “hypertension,” and “suicidal” ideation. This starkly contrasts with the “lifestyle luxury” imagery used in Instagram promotion. The financial extraction is not just a monetary loss; it is experienced as “mental torture.”

```
Game is full of scripted
[events]... For what we play
game's? For Mind relaxation but
this game make you mad and mentally
torture by robbing your wealth.
```

```
This game sucks the life out of
you. It triggers depression because
they up your betting levels even if
you dont win... The developers are
pirates, they will leave you with
nothing.
```

```
Creates suicidal depression if
played.
```

In summary, the Google Play reviews provide a critical empirical counterpoint to the promotional ecosystem on social media. If our analysis of Meta and Instagram reveals how users are *recruited*, through deceptive narratives and AI/deepfake-driven authority, the review analysis reveals how they are *retained* and *exploited* through addictive design and algorithmic bias. Crucially, the qualitative evidence illustrates how these harms compound: a user’s initial financial loss is frequently exacerbated by perceived manipulation, withdrawal friction, and absent customer support, transforming what might have been a discrete negative outcome into a sustained experience of emotional and financial stress. This gap between upstream promises and downstream reality validates the paper’s central claim: that the ecosystem is not merely “risky,” but characterized by a structural disconnect, where aggressive acquisition strategies are inextricably linked to the cascading harms reported after adoption.

6 Discussion

Our mixed-method analysis reveals a digital gambling ecosystem that is fundamentally characterized by a structural disconnect: a widening gap between the “sanitized” promises made during recruitment and the “predatory” reality experienced after adoption. While prior work has examined gambling ads or user harms in isolation, our triangulation of these data sources suggests that they are coupled components of a single, highly optimized ecosystem.

AI as a Force Multiplier. A key finding from our Meta analysis is the industrial scale of deceptive persuasion. The “Easy Money” narrative—present in over 99% of annotated ads—is no longer just a rhetorical strategy; it is a generated artifact. The presence of AI-manipulated media (14.8%) and deepfakes signals a shift in the economics of fraud. Generative AI has lowered the barrier to entry for creating high-quality, authoritative-looking content (e.g., a “CNN-style” news clip or a celebrity endorsement). This technological shift supports a “burn-and-churn” operational playbook: bad actors can rapidly generate synthetic creatives, launch a new Facebook Page, run ads for a few hours to harvest deposits, and abandon the asset before enforcement catches up. This explains the high volume of "removed" content we observed—20,591 ads were only visible as "removed" placeholders in the Meta ad library. While this indicates that moderation systems are eventually triggering, the lag time allows the platform to collect ad revenue while the advertiser successfully recruits victims. The "Playbook" is thus resistant to standard content moderation because the cost of generating new, evasive content has dropped to near zero.

Platform Complicity and the Transparency Gap. The persistence of these ads raises uncomfortable questions about platform incentives. Despite Meta's stated policies prohibiting real-money gambling without prior written permission, our data shows the platform functioning as the primary recruitment engine for an illicit industry. The discrepancy between the "removed" status of ads and the sheer volume of active campaigns suggests a lucrative game of "Whac-A-Mole," where the platform profits from the very evasion tactics it claims to police (Horwitz 2025). Furthermore, we note a critical disparity in transparency. While similar advertisements in the European Union would be subject to strict disclosure requirements regarding the beneficiary and payer, ads targeting Indian users face no such transparency mandates. This regulatory arbitrage allows offshore operators to target users in the Global South with aggressive tactics that would be illegal or immediately flagged in jurisdictions with stricter digital services acts.

The "Shadow Funnel" and Regulatory Evasion. Our qualitative analysis highlights a major challenge for regulation: the Shadow Funnel of recruitment. The primary social platforms (Instagram, Facebook) are rarely the site of the transaction. Instead, they serve as the top of the funnel, utilizing redirection to push users toward encrypted messaging apps (Telegram) or sideloaded APKs. This renders app-store-level enforcement insufficient. Even if Google Play removes a specific betting app, the recruitment infrastructure remains intact on Meta, simply redirecting users to a new domain or a channel on Telegram where tips act as a proxy for unregulated betting. This multi-sited structure explains why our Instagram analysis found a smaller-than-expected organic footprint. Unlike YouTube or Twitch, where live gambling is a content genre (The Hindu 2025), the Instagram ecosystem appears to be a signpost ecosystem where its primary function is not to host the community, but to direct traffic into the dark.

The Human Cost: Interpreting User Reviews. Finally, our analysis of Google Play reviews provides the necessary human counterweight to the "Easy Money" narrative. While we acknowledge the well-documented negativity bias in online reviews (Wu 2013), where dissatisfied users are more likely to post than satisfied ones, the specificity of the complaints in our corpus points to systemic harm rather than varied bad luck. Users do not merely complain about losing; they complain about rigged mechanics ("the algorithm changed after I deposited") and captured funds ("withdrawal pending"). These look like descriptions of fraud, and not of just gambling losses. The consistency of these reports across thousands of reviews validates the use of app store data as a harm sensor. When distinct users use identical language to describe "dynamic difficulty adjustment" after a purchase, it provides strong evidence that the predatory mechanics disguised by the ads are a core feature of the product design.

Limitations. Our study has several limitations. First, our user review dataset is drawn exclusively from the Google Play Store. Since Google maintains stricter policies against real-money gambling than the open web, the apps we analyzed likely represent the cleanest tier of the market. The harms associated with sideloaded apps (APKs) or fully offshore betting sites are likely significantly more severe, as those platforms operate with zero oversight. Second, our annotation protocol prioritized high-level interpretability over subjective nuance. To ensure reliability, we only annotated harms that were explicitly present in the text or visuals (e.g., "guaranteed win"). This conservative approach likely underestimates the prevalence of implicit psychological manipulation or more subtle forms of coercion. Finally, our Instagram analysis was limited to hashtag-based discovery. It is possible that large organic communities exist within private accounts or Stories that are not visible to public hashtag searches.

Future Work. This work suggests two immediate directions for future research. First, we plan to trace the "Shadow Funnel" further by systematically analyzing the Telegram channels and third-party websites linked in these ads. Understanding the conversion scripts used in these private channels is critical to breaking the recruitment chain. Second, a comparative analysis is needed. By expanding this methodology to the European market, we aim to measure the effectiveness of the Digital Services Act and other transparency frameworks. Does the transparency gap lead to measurably different advertising tactics in India and the EU? Answering this question can help policymakers assess whether mandated transparency is enough to deter the coordinated deceptive infrastructure we observed. As a first step, we collected a dataset of approximately 100,000 Meta ads targeting EU users using the same keyword strategy. Although the dataset has not yet been fully cleaned or annotated, our early analysis suggests that this betting app infrastructure operates at substantial scale and is not limited to India. Initial inspection points to similarities in core persuasive narratives, such as easy-money framing and risk minimization, as well as redirection to Telegram and third-party apps. Together, these findings provide preliminary evidence that the underlying recruitment logic may generalize across markets.

Generative AI usage. The authors acknowledge the use of generative AI tools like Gemini for writing assistance in parts of the paper. All AI generated text was double checked by the authors before being included in the paper.

References

Andreou, A.; Silva, M.; Benevenuto, F.; Goga, O.; Loiseau, P.; and Mislove, A. 2019. Measuring the Facebook advertising ecosystem. In *NDSS 2019-Proceedings of the Network and Distributed System Security Symposium*, 1–15.

Atkinson, R.; and Flint, J. 2001. Accessing hidden and hardto-reach populations: Snowball research strategies. *Social research update*, 33(1): 1–4.

Baek, T. H. 2023. Digital advertising in the age of generative AI. *Journal of Current Issues & Research in Advertising*, 44(3): 249–251.

Balhara, Y. P. S.; Gulati, D. D. K.; and Rajguru, A. J. 2024. Fantasy sports as gaming or gambling? Perception, attitudes, and engagement behavior of college students. *Indian Journal of Psychological Medicine*, 46(1): 60–65.

Benegal, V. 2013. Gambling experiences, problems and policy in India: a historical analysis. *Addiction*, 108(12).

Bertaglia, T.; Goanta, C.; Spanakis, G.; and Iamnitchi, A. 2025. Influencer self-disclosure practices on Instagram: A multi-country longitudinal study. *Online Social Networks and Media*, 45: 100298.

Bhatt, A. 2022. Fantasy sports apps are driving a surge in gambling addiction in India. Rest of World. Accessed: 2026-01-07.

Business Standard. 2025. Offshore betting apps use social media, surrogate ads to bypass ban. *Business Standard*. Accessed: 2026-01-14.

Casado-Aranda, L.-A.; Dimoka, A.; and Sánchez-Fernández, J. 2019. Consumer processing of online trust signals: a neuroimaging study. *Journal of Interactive Marketing*, 47(1): 159–180.

Cruickshank, I. J.; and Kloo, I. 2023. Crowdsourcing and Snowball Sampling: Addressing Challenges in Data Collection from Social Media and News Platforms.

Edelson, L.; Lauinger, T.; and McCoy, D. 2020. A security analysis of the Facebook ad library. In *2020 IEEE Symposium on Security and Privacy (SP)*, 661–678. IEEE.

Farid, H. 2021. An overview of perceptual hashing. *Journal of Online Trust and Safety*, 1(1).

Farid, H. 2025. Mitigating the harms of manipulated media: Confronting deepfakes and digital deception. *PNAS nexus*, 4(7): pgaf194.

Financial Express. 2025. Meta tightens rules for real-money gaming advertisers. *The Financial Express*. Accessed: 2026-01-14.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Gandhi, H. 2026. Countries are outlawing online gambling ads. Meta is selling them anyway. Rest of World. Accessed: 2026-01-07.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Gerrard, Y. 2018. Beyond the hashtag: Circumventing content moderation on social media. *New Media & Society*, 20(12): 4492–4511.

Gkiouzepi, E.; Andreou, A.; Goga, O.; and Loiseau, P. 2023. Collaborative ad transparency: Promises and limitations. In *SOSP*. IEEE.

Glaser, B.; and Strauss, A. 2017. *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Hing, N.; Russell, A. M.; Browne, M.; Rockloff, M.; Lole, L.; Tulloch, C.; and Greer, N. 2022. Smartphone betting on sports, esports and daily fantasy sports amongst young adults. *Sydney: NSW Responsible Gambling Fund*.

Hing, N.; Vitartas, P.; Lamont, M.; and Fink, E. 2014. Adolescent exposure to gambling promotions during televised sport: An exploratory study of links with gambling intentions. *International Gambling Studies*, 14(3): 374–393.

Hoonka, A. 2022. An Analysis of Sports Betting in India. *Part 2 Indian J. Integrated Rsch. L.*, 2: 1.

Horwitz, J. 2025. Meta is earning a fortune on a deluge of fraudulent ads, documents show. Reuters. Accessed: 2026-01-07.

Houghton, S.; McNeil, A.; Hogg, M.; and Moss, M. 2019. Comparing the Twitter posting of British gambling operators and gambling affiliates: A summative content analysis. *International Gambling Studies*, 19(2): 312–326.

Hwang, S.; Nanayakkara, P.; and Shvartzshnaider, Y. 2025. Trust and friction: Negotiating how information flows through decentralized social media. *Proceedings of the ACM on Human-Computer Interaction*, 9(7): 1–26.

India Today. 2025. Real money gambling: Addiction, debt, fraud, and a nation trapped. *India Today*. Originally published in the India Today issue dated June 2, 2025.

Krishnamoorthy, R.; Anandharaja, E.; Ajitha, S.; and Muralidhar, P. 2025. Online Gambling Preferences and Consequences. In *Applied Research for Growth, Innovation and Sustainable Impact*, 370–376. Routledge.

Krishnan, M. 2025. India: Addicted to online betting and fantasy gaming. DW. Accessed: 2026-01-07.

Le Pochat, V.; Edelson, L.; Van Goethem, T.; Joosen, W.; McCoy, D.; and Lauinger, T. 2022. An audit of Facebook's political ad policy enforcement. In *31st USENIX Security Symposium (USENIX Security 22)*, 607–624.

McGrane, E.; Wardle, H.; Clowes, M.; Blank, L.; Pryce, R.; Field, M.; Sharpe, C.; and Goyder, E. 2023. What is the evidence that advertising policies could have an impact on gambling-related harms? A systematic umbrella review of the literature. *Public Health*, 215: 124–130.

Pitt, H.; McCarthy, S.; Randle, M.; Daube, M.; and Thomas, S. L. 2024. Young people's views about the use of celebrities and social media influencers in gambling marketing. *Health Promotion International*, 39(1): daae012.

Singer, J.; and Wöhr, A. 2024. Gambling operators' use of advertising strategies on social media and their effects: A systematic review. *Current Addiction Reports*, 11(3).

Singh, J. 2025. India bans real-money gaming, threatening a $23 billion industry. Accessed: 2026-01-14.

Stadder, E.; and Naraine, M. L. 2020. Place your bets: An exploratory study of sports-gambling operators' use of twitter for relationship marketing. *International Journal of Sport Communication*, 13(2): 157–180.

The Hindu. 2025. Actors, anchors, TV hosts among 29 booked by ED for promoting illegal betting apps. *The Hindu*. National, Telangana section; online article.

Tiwari, A. K. 2023. In India, fantasy gaming is causing addiction and financial ruin. *Al Jazeera*. Accessed: 2026-01-14.

Wardle, H.; Degenhardt, L.; Marionneau, V.; Reith, G.; Livingstone, C.; Sparrow, M.; Tran, L. T.; Biggar, B.; Bunn, C.; Farrell, M.; et al. 2024. The lancet public health commission on gambling. *The Lancet Public Health*, 9(11): e950–e994.

Wu, P. F. 2013. In search of negativity bias: An empirical study of perceived helpfulness of online reviews. *Psychology & Marketing*, 30(11): 971–984.

Xiao, L. Y. 2025. Illegal loot box advertising on social media? An empirical study using the Meta and TikTok ad transparency repositories. *Computer Law & Security Review*.

Ethics checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, we have listed our data collection methodology and follow all privacy norms.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, claims made in the abstract and introduction are supported by our results in Section 5, with more detailed experimental setup and results also listed in the Appendix
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, we review prior methods in Section 2, and in Sections 3 and 4 we outline our data collection and annotation procedures. Together, these sections clarify the rationale behind our proposed methodology.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, we list this in Section 3 on data collection
   (e) Did you describe the limitations of your work? Yes, we list them in Section 6
   (f) Did you discuss any potential negative societal impacts of your work? NA
   (g) Did you discuss any potential misuse of your work? NA
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA, we do not have theoretical results.
   (b) Have you provided justifications for all theoretical results? NA, we do not have theoretical results.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, we have listed about the data in Section 3, instructions and experimental results in Section 4, 5
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes, we ran the experiments on Tesla P100 16GB GPU provided by Kaggle Notebooks.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes
   (b) Did you mention the license of the assets? No
   (c) Did you include any new assets in the supplemental material or as a URL? Yes
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? NA

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA

(d) Did you discuss how data is stored, shared, and deidentified? NA

A Appendix

A.1 Dataset Statistics

The following keywords were used for data collection: betM sports, Match Book App, Stake game, crickex, 20Bet, Spin Casino, crypto trading, online betting, mahadev betting app, 3 Lucky Casino, World777, Indibet, Teen Patti Kash, Betwinner, Sportsbet.io, Kheloyar, grand mondial app, cricket betting, gofun88, sports betting, Jackpot City Casino, Bambet, Puntit, 188bet, Megapari, 91Club, Blackjack Live, 96p.in, 3patti Guide, Teen Patti Romio, Dafabet, Aviator Game, Live Baccarat, betting sites, Betway india, Fun88, 1win, BetBook247, Fairplay, Binomo, Betting United: Betting Tips, 10Cric, 4rabet, bc.game, 22bet, betkwiff, Wheel of Fortune betting, IPL betting, Oppa 888, Mostbet, Gold365 app, Melbet, Tivit Bet App, Aviator, color prediction trading, JeetWin, AndarBahar Poker, IPL win prediction, SkyExchange, crypto betting, Rajabets, color trading, Fairplay 24, Lotus365, Teen Patti Lucky Gold, 1XBet, Zet Casino, Stake.com, betongame india, IPL score prediction, Odds96, Roobet, IPL fantasy betting, Krundi, Betwaysatta, Parimatch, Wolf777, Casumo Online Games, PokerStars, jeetbuzz, Mahadev Book, Betindi, color prediction betting, Bet365

Ad Category	Count
Fiction story application ads	290,394
Ads taken down by Meta at the time of data collection	20,591

Table 3: Summary of unrelated content and unavailable ads in the Meta Ad Library dataset

A.2 Meta Ads Annotation

For each advertisement, annotators were asked to answer the following questions:

1. **Content relevance identification:** Does this advertisement appear to be related to betting/ gambling?
   - Yes
   - No
2. **Ad category:** Which category best describes the primary product or service being advertised? (Select all that apply.)
   - Sports betting
   - Fantasy sports
   - Casino games
   - Card games
   - Prediction games
   - Lottery / jackpots
   - Trading / cryptocurrency
   - Other (please specify)
3. **App / platform name:** Which application or platform is being promoted? (Select all that apply.)
   - List of names such as Parimatch, Betway, 1xBet, 1win, 22bet, Bet365, BC Game etc.
   - Other (please specify)
4. **Primary messaging strategy:** What is the dominant messaging or persuasive strategy used in the advertisement? (Select all that apply.)
   - User acquisition
   - Promoting a new feature
   - Celebrity endorsement
   - Misleading or deceptive claims
   - Emotional appeal
   - Social proof/ testimonials
   - General promotion/ brand visibility
5. **Potentially harmful narratives:** Which potentially harmful narratives are present in the advertisement? (Select all that apply.)
   - Easy Money Narrative
   - Risk Minimization
   - Income Source Framing
   - Addiction Normalization
   - Debt Solution Marketing
   - Success Story Manipulation
   - Time Pressure Tactics
   - Other (please specify)
6. **Media Authenticity:** How would you characterize the media content in the ad?
   - Authentic
   - AI-generated
   - Deepfake

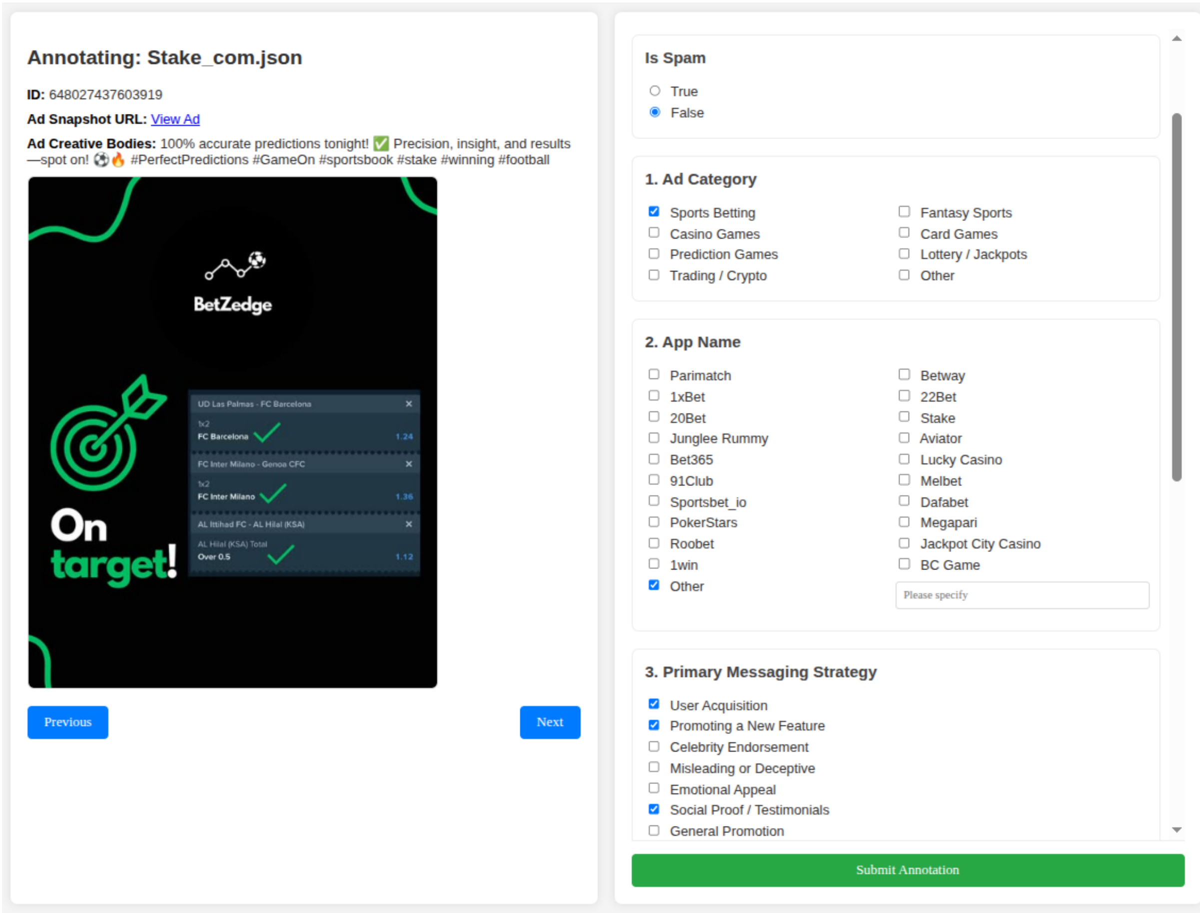


Figure 2: Meta Ads Annotation UI

7. **Presence of sexualized content:** Does the advertisement contain sexualized imagery or themes?
   - Yes
   - No
8. **Additional notes**
   - text field for any observations or contextual remarks.

A.3 Few-shot Classification using Gemini

We use `gemini-2.5.-pro` to perform annotation's of Instagram Posts via few-shot classification. e model was provided with a small number of labeled examples (images with annotations) followed by a main prompt requesting classification of a new image. Below is the prompt we provided.

```
Analyze this image and answer the
    following questions. Provide your
     response in JSON format.

IMPORTANT: First determine if the
    image is unrelated or not. For
    the image to be unrealted the
    content should NOT be related to
    gambling, betting, or similar
    activities.
If the image is unrealted, then you
    don't have to answer the other
    questions.

1. Is unrealted: Is this image
    unrealted content? (true/false)

IF AND ONLY IF "is_unrealted" is
    FALSE, answer the following
    questions. Otherwise, skip them:

2. Ad Category: Select all that
    apply from:
   - Sports Betting
   - Fantasy Sports
   - Casino Games
   - Card Games
   - Prediction Games
   - Lottery / Jackpots
   - Trading / Crypto
   - Other (specify)
```

```
3. App Name: Identify any of these
    apps mentioned or shown:
   - Parimatch, Betway, 1xBet, 22Bet
       , 20Bet, Stake, Junglee Rummy,
        Aviator, Bet365,
     Lucky Casino, 91Club, Melbet,
         Sportsbet_io, Dafabet,
         PokerStars, Megapari,
     Roobet, Jackpot City Casino, 1
         win, BC Game
   - Other (specify)

4. Primary Messaging Strategy:
    Select all that apply:
   - User Acquisition
   - Promoting a New Feature
   - Celebrity Endorsement
   - Misleading or Deceptive
   - Emotional Appeal
   - Social Proof / Testimonials
   - General Promotion

5. Potentially Harmful Narratives:
    Select all that apply:
   - Easy Money Narrative
   - Risk Minimization
   - Income Source Framing
   - Addiction Normalization
   - Debt Solution Marketing
   - Success Story Manipulation
   - Time Pressure Tactics
   - Other (specify)

6. Media Authenticity & Manipulation
    : Select all that apply:
   - Authentic
   - AI-Generated (Synthetic Media)
   - Deepfake (Celebrity)
   - Deepfake (Non-Celebrity)

7. Sexual Content: Does the image
    contain sexual content? (yes/no)

8. Ad Notes: Write 1-2 sentences
    describing the image. Write what
    you think are harmful things in
    this image for the user. Write
    any additional information about
    the image if useful.

---
Here are some examples to guide you:
<list of examples>
---

Now analyze the NEW IMAGE below and
    provide your annotation.
```

```
Respond ONLY with a valid JSON
    object in this format:

If is_unrealted is TRUE:
{
  "is_unrealted": true
}

If is_unrealted is FALSE:
{
  "is_unrealted": false,
  "ad_category": ["category1", "
      category2"],
  "app_name": ["app1", "app2"],
  "primary_messaging_strategy": ["
      strategy1"],
  "potentially_harmful_narratives":
      ["narrative1"],
  "media_authenticity": ["type1"],
  "sexual_content": "yes"/"no",
  "ad_notes": "text"
}
```

Below, we list the few-shot image examples provided to gemini along with their annotations.

Example 1

Content Classification: Unrelated content (not related to gambling or betting).

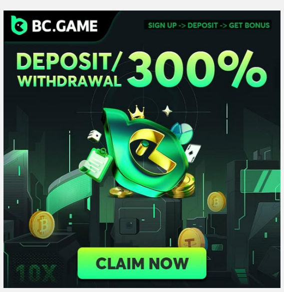
Example 2

Content Classification: Unrelated content (not related to gambling or betting).

Example 3

Content Classification: Unrelated content (not related to gambling or betting).

Example 4



Content Classification: Related content
Ad Category: Casino Games; Trading/Crypto
App Name: BC Game
Messaging Strategy: User Acquisition; General Promotion
Potentially Harmful Narratives: Easy Money Narrative; Risk Minimization
Media Authenticity: Authentic
Sexual Content: No
Ad Notes: The ad promotes BC.Game, an online casino, with a 300% deposit bonus. This framing emphasizes an easy-money narrative and minimizes perceived risk.

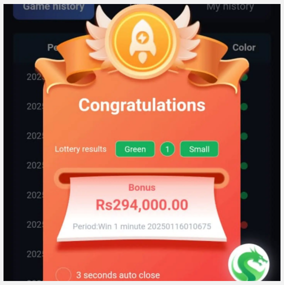
Example 5

Content Classification: Related content
Ad Category: Sports Betting
App Name: Betway; 1xBet; Parimatch; 1win; Aviator
Messaging Strategy: User Acquisition; General Promotion
Potentially Harmful Narratives: Easy Money Narrative; Risk Minimization
Media Authenticity: Authentic
Sexual Content: No
Ad Notes: The image promotes an ’Aviator Predictor’ app version 6.0 and various betting platforms.

Example 6

Content Classification: Related content
Ad Category: Sports Betting; Prediction Games
App Name: Betway; 1xBet; Parimatch; 1win; Aviator
Messaging Strategy: User Acquisition; General Promotion
Potentially Harmful Narratives: Easy Money Narrative; Risk Minimization
Media Authenticity: Authentic
Sexual Content: No
Ad Notes: The image promotes an ’Aviator Predictor’ app version 6.0 and various betting platforms. The easy money narrative and potential misleading claims are harmful.

Example 7



Content Classification: Related content
Ad Category: Lottery/Jackpots; Prediction Games
App Name: None specified
Messaging Strategy: Misleading or Deceptive; General Promotion
Potentially Harmful Narratives: Easy Money Narrative
Media Authenticity: Authentic
Sexual Content: No
Ad Notes: The image shows a lottery game winning announcement with a bonus of Rs294,000. It promotes an ’easy money’ narrative.

Online Gambling and Games

Meta defines online gambling and games as any product or service where anything of monetary value is included as part of a method of entry and prize. Ads that promote online gambling and gaming are only allowed with our prior written permission. Authorised advertisers must follow all applicable laws and include targeting criteria consistent with Meta's targeting requirements. At a minimum, ads may not be targeted to people under 18 years of age. Learn more in our Business Help Centre.

Learn more

(a)

Gambling fraud and scams

Content that:

- Offers real money gambling services ("Real money" is real-world currency that can be used to buy goods or services in the real world, including national currencies such as US Dollars and virtual currencies such as Bitcoin):
  - with a guarantee of winning.
  - implying or admitting to have rigged the outcome of a game or match.
  - soliciting people to enable match fixing or looking for help or tips on how to fix a match or game.

Social casino games that simulate gambling with no opportunity to win real money fall under our Community Standard for Restricted Goods and Services.

(b)

Figure 3: Screenshot of Meta Policy that lists Online betting and gambling content is not allowed. These screenshots are taken from https://transparency.meta.com/policies/ad-standards/ and https://transparency.meta.com/policies/community-standards/fraud-and-scams/