ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 6| June 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

ALGORITHMIC AMPLIFICATION IN ADULT MEDIA PLATFORMS: A STUDY ON RECOMMENDATION OF SENSITIVE CONTENT

Abhijit Joshi[1], Soumya Ranjan Bhoi[2], Nikhil Agrawal[3]
[1]*Assistant Professor, Dept. of Computer Science, Rajendra University, Balangir,*
[2]*Assistant Professor, Dept. of Computer Science, Rajendra University, Balangir,*
[3]*Assistant Professor, Dept. of Philosophy, Maa Manikeshwari University, Bhawanipatna*

Article DOI: https://doi.org/10.36713/epra28279
DOI No: 10.36713/epra28279

ABSTRACT

Adult media platforms remain underexamined in recommender-systems and algorithmic fairness research despite their scale, intimate data environment, and reliance on personalization. This paper develops a socio-technical framework for studying algorithmic amplification in adult-platform recommendation systems, focusing on whether sensitive content is surfaced at rates above a user's baseline exposure. Building on warning-amplification logic from sensitivity-aware recommendation research, the paper formalizes amplification through prevalence-based and normalized measures that compare recommendation outputs with user histories. It combines these measures with metadata analysis, semantic clustering of tags, comparative benchmarking against general-media datasets, and persona-based black-box auditing. The paper argues that adult-platform recommenders should be treated as a distinct high-risk domain because sensitive categories are central to content organization rather than peripheral. It concludes with an ethical protocol and a roadmap for future empirical audits under GDPR- and DSA-relevant governance concerns.

KEYWORDS: *Algorithmic amplification, adult media platforms, recommender systems, warning amplification, sensitive content*

1. INTRODUCTION

Recommender systems are at the core of the modern platform economy because they decide what users see, in what order, and under what forms of personalization. These systems do more than organize content; they shape attention, visibility, and user experience through candidate generation, ranking, and feedback-driven optimization, and their broader societal effects have already been studied in mainstream social media and entertainment contexts in relation to polarization, misinformation, and other forms of harmful amplification.

Yet adult media platforms remain comparatively underexamined despite their scale, commercial significance, and dependence on similar algorithmic infrastructures. Prior work on Pornhub has shown that recommendations vary according to inferred or self-reported user attributes such as gender and sexual interest, while sensitivity-aware benchmark datasets in general-media recommendation have shown that user exposure to harmful or warning-bearing material can be formally evaluated beyond traditional accuracy metrics. Together, these developments provide a strong basis for asking whether recommenders for adult media amplify sensitive, stereotyped, or aggressive content differently from recommenders in general-media environments.

This paper investigates that question by examining recommendation behavior on pornographic media platforms through the lens of algorithmic amplification. More specifically, it asks whether recommendation outputs deviate from the baseline patterns present in a user's interaction history and whether such outputs disproportionately expose material associated with aggression, rigid gender norms, racialized descriptors, or other sensitive classifications. The study is motivated by the fact that adult platforms combine highly personalized recommendation with extremely sensitive behavioral data, creating a socio-technical environment in which algorithmic effects may carry stronger psychological, social, and regulatory implications than in ordinary entertainment domains.

To make this inquiry analytically tractable, the paper adopts the notion of warning amplification and extends it to adult-platform settings where content sensitivity is often encoded through platform metadata rather than through standardized warning ontologies. It also draws on prior algorithmic audits and metadata studies showing that taxonomies, tags, interface design, and recommendation systems interact to shape what sexual content becomes visible and how identities and practices are represented. The aim is therefore not merely to criticize adult-platform recommendation morally, but to provide a formal and interdisciplinary framework through which sensitive-content amplification can be studied, compared, and eventually audited in a technically reproducible way.

The rest of the paper is organized as follows. Section 2 reviews related work on recommender systems, adult-platform studies, metadata, and audit methods. Section 3 develops the theoretical framework and formalizes the amplification metric. Section 4 presents the research questions and hypotheses. Section 5 sets out the methodology, including data sources, operationalization, semantic grouping, and persona-based auditing. Section 6 offers a socio-technical discussion of the framework's implications. Section 7 outlines limitations and the ethical protocol that should govern future empirical use of the

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 6| June 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

framework. Section 8 provides a roadmap for empirical implementation. Section 9 concludes by positioning the framework as a bridge between technical auditing and platform governance under emerging regulatory regimes such as the Digital Services Act.

2. RELATED WORK

Research on recommender systems has established that algorithmic curation is not a neutral delivery mechanism, but a tool that shapes visibility, attention, and user exposure through ranking and personalization. In mainstream digital environments, recommendation pipelines have been associated with amplification dynamics linked to engagement optimization, polarization, and other forms of systemic risk, which has led to broader concern about algorithmic governance and platform accountability.

A second line of work has focused on sensitivity-aware recommendation evaluation. Datasets such as ML-DDD and AO3 associate user-preference data with content warnings and allow researchers to study whether recommenders surface harmful or warning-bearing items at rates that exceed or fall below a user's historical baseline, thereby extending evaluation beyond accuracy-only paradigms. This literature is especially important for the present paper because it provides the conceptual foundation for warning amplification as a measurable exposure effect.

A third body of scholarship examines the platformization of adult media and the algorithmic organization of sexuality. Research on Pornhub and related platforms has shown that recommendation outputs can vary widely according to inferred or declared user identity, gender, and sexual interest, suggesting that personalization is entangled with demographic profiling and normative assumptions about sexuality rather than operating as a purely neutral matching process. This literature also argues that adult-platform taxonomies and tagging systems do not merely reflect user demand but help construct commercially preferred and culturally patterned portrayals of desire, identity, and bodily representation.

Related scholarship has also found that adult-platform recommendation can reinforce stereotypes in cold-start settings, homepage ranking, and search suggestions. Audit studies using scripted browsing sessions have shown that default recommendation flows can overrepresent certain bodies, identities, or performance styles and can associate some demographic markers with more aggressive or sensational descriptions, indicating that bias may originate not only in user histories but also in interface design, default ranking, and metadata organization.

Another relevant line of work concerns metadata, folksonomies, and representation learning. Prior analyses show that platform-controlled taxonomies and user-generated tags tend to converge in highly visible clusters, creating semantic structures that can be learned by recommendation models and embedding systems; studies using models such as MiniLM have further shown that adult-platform tag spaces can encode stereotyped associations later inherited by downstream retrieval and ranking systems. This is directly relevant here because it demonstrates that recommendation bias may be embedded not only in ranking logic but in the representational space on which recommendation operates.

Finally, the literature on algorithmic auditing provides methodological tools for studying opaque systems through black-box observation, UI scraping, persona simulation, and scenario-based hypothesis testing. Where commercial platforms do not provide meaningful API access or model transparency, such methods allow researchers to infer recommendation behavior from observable outputs rather than proprietary internals, making them especially suitable for adult-platform environments. Taken together, these strands of work motivate the present paper's central claim that adult-platform recommender systems should be studied as a distinct and underexamined case of algorithmic amplification.

3.THEORETICAL FRAMEWORK

The present study begins from the premise that recommender systems are not passive content-delivery tools but optimization systems that shape visibility according to platform-specific objectives. This is particularly significant in adult media, where ranking and personalization may not simply mirror user preference but may also intensify particular categories of content through feedback loops, repeated exposure, and metadata-driven clustering. The central theoretical question is therefore whether sensitive material on pornographic platforms is preserved, suppressed, or amplified relative to the user's observed or constructed baseline.

To study this question, the paper adopts the concept of algorithmic amplification, understood here as the disproportionate visibility of certain content features beyond what would be expected from organic user behavior alone. In this setting, amplification is not a metaphor but a measurable structural effect of recommendation pipelines. The relevant contrast is between what appears in a user's baseline history and what is surfaced by the platform in ranked or recommended output, since that comparison allows one to distinguish simple repetition from selective elevation.

3.1 Warning amplification

The most direct formal basis for this comparison comes from warning amplification, a metric used in sensitivity-aware recommendation research to compare the prevalence of warning-bearing items in recommendation outputs with their prevalence in a user's relevant history. In the present paper, that logic is adapted to adult-platform settings, where sensitive categories are often identified through metadata, tags, titles, or coded descriptors rather than through a universally standardized warning ontology. For a user or audit session u and a sensitive category c, let H_u denote the relevant baseline history and R_u the set of recommended items observed for that user or session. Let $L(i)$ denote the set of labels, warnings, tags, or coded metadata attached to item i. The prevalence of category c in the baseline history is defined as

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 6| June 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

$$p_H(u,c) = \frac{1}{|H_u|} \sum_{i \in H_u} \mathbf{1}\{c \in L(i)\} \quad (1)$$

and the prevalence of the same category in the recommendation output is defined as

$$p_R(u,c) = \frac{1}{|R_u|} \sum_{j \in R_u} \mathbf{1}\{c \in L(j)\}. \quad (2)$$

Here, $\mathbf{1}\{\cdot\}$ is an indicator function equal to 1 when the item carries category c and 0 otherwise. The primary amplification score used in this paper is then

$$A(u,c) = p_R(u,c) - p_H(u,c). \quad (3)$$

Under this formulation, $A(u,c) > 0$ indicates amplification, $A(u,c) = 0$ indicates parity, and $A(u,c) < 0$ indicates deamplification. This difference-based formulation preserves the logic of the warning-amplification literature while making the metric more suitable for black-box auditing conditions in which baseline exposure is heterogeneous and category coding is metadata-driven rather than ontology-given.

"Relevant history" depends on the interaction regime. In explicit-rating scenarios, the baseline is often derived from positively rated items, sometimes restricted to the top quartile of ratings, while in implicit-feedback environments any positive action such as a click, like, bookmark, or sustained engagement may serve as evidence of endorsement. This distinction matters because adult platforms often rely on implicit behavioral traces rather than explicit preference declarations, so the theoretical framework must remain compatible with both explicit and implicit settings.

3.2 General-media baseline

Prior work using this framework on general-media datasets suggests that personalized recommendation does not always amplify sensitive material and may sometimes deamplify it relative to a user's baseline. Studies on ML-DDD and AO3 indicate that collaborative-filtering models such as SVD and ALS can recommend fewer warning-bearing items than are present in a user's historical preference set, whereas non-personalized baselines such as Random or TopPop can show stronger amplification because they track global popularity rather than individualized structure. This is theoretically important because it suggests that personalization can moderate exposure when sensitive content is relatively sparse or peripheral to the broader corpus.

In the present paper, that result is treated as a contrast case rather than a general rule. If sensitive content is marginal in a general-media setting, latent-factor models may smooth it out through averaging and dimensional reduction, but adult media platforms differ structurally because sensitive, aggressive, and highly specific content may be central rather than exceptional. The theoretical expectation is therefore not a simple transfer of the general-media pattern, but the possibility of a domain-specific reversal.

3.3 Adult-platform structure

Adult platforms are structurally distinct in that their content is organized around sexuality, identity, body type, performance style, and other sensitive categories. They are not merely content libraries but classification systems in which metadata, interface design, and recommender logic interact to determine what becomes visible and profitable. Prior audits of Pornhub suggest that recommendation behavior can vary by inferred or declared user gender and sexual interest, implying that the system segments users and reinforces particular normative patterns in ways that differ from mainstream media environments.

The framework therefore treats platform taxonomies and user-generated folksonomies as interdependent semantic layers. Official categories are top-down structures for organizing inventory, while user tags are bottom-up descriptions of content and preference; over time, these layers may converge around highly visible content clusters, giving platform-defined categories an appearance of naturalness and inevitability. This matters theoretically because recommendation systems do not operate over raw content alone but over a representational space that has already been socially and commercially shaped.

3.4 Semantic representation

A further part of the framework concerns the encoding of platform metadata through language models and embedding systems. Such models place category labels, titles, and tags into vector spaces in which similarity becomes a computational relation, meaning that recommendation outputs may inherit assumptions embedded in the labeling system rather than discover relations in a neutral manner. In adult-platform environments, this raises the possibility that observed amplification is not solely a function of user history but also of the structure of the metadata space itself.

Let each label or tag t be represented by an embedding vector $\mathbf{e}_t \in \mathbb{R}^d$. Semantic similarity between two labels t_a and t_b can be measured using cosine similarity:

$$\mathrm{sim}(t_a, t_b) = \frac{\mathbf{e}_{t_a}^{\top} \mathbf{e}_{t_b}}{\|\mathbf{e}_{t_a}\| \|\mathbf{e}_{t_b}\|}. \quad (4)$$

A semantic cluster can then be formed when $\mathrm{sim}(t_a, t_b) \geq \tau$, where τ is a researcher-defined threshold. This grouping reduces fragmentation caused by synonymous or near-synonymous descriptors and allows amplification to be estimated over semantic classes rather than isolated lexical forms. The framework thus treats amplification as a function of three interacting layers: user behavior, platform taxonomy, and algorithmic representation.

3.5 Working hypothesis

From this perspective, the paper advances a working hypothesis: if adult platforms are built around sensitive content as a core commercial product, their recommendation systems may be more likely than general-media recommenders to amplify stereotyped, aggressive, or otherwise sensitive material. This is not assumed

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 6| June 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

as a deterministic fact but proposed as the guiding hypothesis for the framework. The rest of the paper therefore uses warning amplification, metadata structure, and platformization as the key conceptual tools for assessing whether adult-platform personalization deviates from the deamplification tendencies sometimes observed in mainstream recommendation contexts.

4. RESEARCH QUESTIONS AND HYPOTHESES

The overarching research question is whether recommender systems on adult media platforms shape exposure to sensitive content differently from recommender systems in mainstream digital environments. Prior work on general-media datasets indicates that personalized models can deamplify sensitive content relative to user history, while audit studies of Pornhub suggest that adult-platform recommendation and interface structures may reinforce gendered and heteronormative patterns in content delivery. This contrast motivates the following research questions and hypotheses.

4.1 Research questions

RQ1. Do recommender systems for adult media domains tend to amplify sensitive content relative to a user's baseline interaction history, rather than simply reproducing or deamplifying it as seen in some general-media settings?

RQ2. Are the level and type of amplification different across user conditions, especially between anonymous cold-start sessions and profile-based or persona-driven sessions?

RQ3. Are certain types of sensitive content, such as aggressive descriptors, racialized labels, or stereotyped gender classifications, more likely than others to be surfaced by adult-platform recommendation systems?

RQ4. How much do platform taxonomies, user-generated tags, and metadata embeddings contribute to the amplification of sensitive categories through the shaping of the semantic structure on which recommendation systems operate?

RQ5. How do amplification patterns on adult platforms differ from those observed in sensitivity-aware benchmark datasets such as ML-DDD and AO3?

4.2 Hypotheses

H1. Adult-platform recommender systems will show greater sensitive-content amplification than general-media recommendation systems measured under comparable warning-amplification logic.

H2. Cold-start recommendation outputs on adult platforms will already exhibit structured bias toward some demographic and stereotyped content categories, indicating that amplification is not only a function of long interaction histories but is partly built into platform defaults.

H3. Profile- or persona-conditioned sessions will produce stronger amplification of narrow, gendered, racialized, or aggressive content categories than anonymous sessions because personalization mechanisms reinforce high-confidence inferences about user preference.

H4. Categories that are strongly encoded in platform metadata, such as official taxonomies and recurring user tags, will be more likely to be amplified than categories that are less semantically stabilized in the platform's classificatory system.

H5. Adult-platform recommendation behavior will deviate from the deamplification tendencies observed in collaborative-filtering evaluations on ML-DDD and AO3, suggesting that content-domain structure materially influences amplification outcomes.

4.3 Scope of hypothesis testing

These hypotheses guide an audit-based and comparative analysis rather than presupposing deterministic causal conclusions. The framework does not assume that all adult-platform recommendation systems amplify all types of sensitive content equally; rather, it asks whether amplification occurs systematically under identifiable conditions of platform design, metadata structure, and user profiling. This is important because it keeps the paper analytical and prevents overstatement in the absence of direct access to proprietary model internals.

5. METHODOLOGY

This study employs a comparative black-box audit approach to examine whether recommender systems on adult media platforms amplify sensitive content differently from recommenders in general-media contexts. The design is driven by a practical constraint: commercial adult platforms do not disclose their recommendation pipelines, ranking features, or personalization logic in a form suitable for direct scientific inspection. The study therefore infers recommendation behavior from publicly available metadata, recommendation-output traces, and sensitivity-labeled benchmark datasets rather than from privileged access to proprietary systems.

5.1 Research design

The research design combines two complementary strategies: comparative benchmarking and platform auditing. Comparative benchmarking uses datasets such as ML-DDD and AO3 to match user-preference data with warning labels and thereby establish amplification patterns in general-media recommendation settings, while platform auditing examines adult-platform behavior through interface-level observation, metadata analysis, and recommendation traces from persona-based sessions. Together, these strategies make it possible to assess the extent to which adult-platform recommendation behavior deviates from the deamplification patterns documented in sensitivity-aware mainstream benchmarks.

The study proceeds at three analytical levels. The first examines the semantic structure of platform content through titles, categories, and tags; the second studies recommendation outputs under different browsing conditions such as anonymous and persona-conditioned sessions; and the third compares those adult-platform patterns with amplification results from general-media benchmarks.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 6| June 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

5.2 Data sources

The adult-platform portion of the study draws on two kinds of material: large-scale content metadata and interface-level recommendation traces. For metadata analysis, the paper uses the Nikity/Pornhub dataset, a public multilingual dataset of video titles and URLs suited to lexical, semantic, and categorical analysis. For interface-level evidence, the study relies on repositories and prior audit traces associated with work such as Tracking Exposed and related Pornhub audits that recorded homepage variation, suggested videos, and profile-dependent recommendation differences across simulated accounts.

The comparative portion uses the sensitivity-aware general-media datasets ML-DDD and AO3. These datasets are methodologically useful because they link user-preference data to explicit warning taxonomies and permit direct calculation of amplification scores in controlled recommendation settings, though they are treated here as benchmarks rather than as domain matches to adult content.

5.3 Unit of analysis

The main unit of analysis is the recommended item shown to a user or simulated persona in a recommendation context such as a homepage carousel, sidebar, or search-adjacent suggestion list. Each item is analyzed with respect to the metadata or warning categories it carries and the user history or browsing condition under which it is surfaced, allowing amplification to be measured as the difference between baseline prevalence and recommendation prevalence rather than as a vague impression of platform bias.

A second unit of analysis is the browsing session itself, especially under cold-start and persona-based audit conditions. Session-level analysis is necessary because adult-platform recommendation behavior may vary not only by item characteristics but also by interaction sequence, account state, localization, and inferred user profile.

5.4 Operationalization of sensitive content and amplification

The principal outcome variable is the amplification of sensitive content. In benchmark datasets such as ML-DDD and AO3, sensitive content is operationalized through explicit warning labels attached to items, whereas in the adult-platform setting it is operationalized through observable metadata classes derived from titles, tags, platform categories, and visible descriptors. These classes may include aggressive descriptors, stereotyped gender labels, racialized categories, and other markers that recur within the platform's metadata ecology.

Because users and sessions may begin from very different baseline exposure levels, the study also defines a normalized amplification score:

$$A_{\text{norm}}(u,c) = \frac{p_R(u,c) - p_H(u,c)}{\max\{p_H(u,c), \varepsilon\}}, \varepsilon > 0. \quad (5)$$

The constant ε prevents division by zero in cold-start or sparse-history conditions, which are central to the audit design. This normalized score is useful when comparing anonymous sessions, lightly conditioned personas, and mature interaction histories within a common framework.

The crucial point is that adult-platform data do not come with a universally standardized warning ontology. The coding of aggression, stereotype, demographic targeting, or related forms of sensitivity must therefore be treated as an explicit operational decision rather than a naturally given taxonomy. This does not weaken the methodology; instead, it makes the interpretive layer of the analysis visible and therefore contestable.

5.5 Baseline and comparison conditions

For general-media datasets, the baseline is the user's relevant interaction history, following the logic of sensitivity-aware recommendation evaluation. In explicit-feedback settings this baseline is typically built from highly rated items, while in implicit-feedback settings it is constructed from positive interactions such as likes, bookmarks, or kudos. This provides a structured reference point against which recommendation outputs can be compared.

In the adult-platform audit setting, two baseline conditions are central. The first is the cold-start baseline, in which a clean browser or new session with no prior interaction history is used to observe what the platform surfaces by default. The second is the persona-history baseline, in which simulated users repeatedly interact with a narrow class of content and the resulting recommendation outputs are observed over time. These baselines make it possible to distinguish default platform bias from profile-conditioned amplification.

5.6 Analytical method

The analysis proceeds in four stages. First, adult-platform metadata are collected or retrieved from public repositories and cleaned through category normalization, lexical deduplication, and grouping of semantically related descriptors. Second, recommendation traces from anonymous and persona-based sessions are organized by recommendation surface, session type, and profile condition. Third, the prevalence of sensitive categories is calculated separately for baseline histories and recommendation outputs. Fourth, amplification scores are computed and compared across session types, metadata classes, and benchmark datasets.

At the aggregate level, category-specific amplification across a set of users or audit sessions U may be defined as

$$\bar{A}(c) = \frac{1}{|U|} \sum_{u \in U} A(u,c). \quad (6)$$

Where users or sessions contribute substantially different numbers of observations, a weighted version may also be used:

$$\bar{A}_w(c) = \frac{\sum_{u \in U} w_u A(u,c)}{\sum_{u \in U} w_u}. \quad (7)$$

These measures allow the analysis to move from item-level or session-level comparisons to platform-level patterns, which is necessary for the paper's broader domain claims.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 6| June 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

5.7 Persona-based audit approach

To address RQ2 and RQ3, the framework employs a persona-based audit design inspired by prior Pornhub auditing work. Simulated profiles are assigned distinct browsing patterns or identity declarations, and their recommendation outputs are monitored over repeated interactions. Some of these profiles function as controls with no meaningful history, while others are designed to focus narrowly on selected semantic clusters so that downstream recommendation shifts can be observed.

This approach does not assume that persona simulation perfectly reproduces lived user experience. Rather, it functions as an experimental probe of platform behavior under controlled and repeatable conditions. Differential amplification between cold-start and persona-conditioned sessions may be defined as

$$\Delta_{\text{persona}}(c) = \frac{1}{|S_1|}\sum_{s\in S_1} A(s,c) - \frac{1}{|S_0|}\sum_{s\in S_0} A(s,c). \qquad (8)$$

where S_0 denotes cold-start sessions and S_1 persona-conditioned sessions. If $\Delta_{\text{persona}}(c) > 0$, persona conditioning increases amplification relative to default exposure; if $\Delta_{\text{persona}}(c) < 0$, personalization dampens the category relative to cold-start outputs.

5.8 Comparative interpretation

The adult-platform results are interpreted in relation to benchmark results from general-media settings rather than in isolation. If general-media personalized recommenders tend to smooth or deamplify warning-heavy content while adult-platform outputs tend to elevate aggressive or stereotyped categories, that contrast would support the claim that domain structure materially affects amplification behavior. The comparative design thus enables the paper to move beyond moral critique toward a formally grounded analysis of recommendation across content domains.

6. SOCIO-TECHNICAL DISCUSSION

The central implication of this paper is that recommender systems cannot be evaluated independently of the domains in which they operate. While prior work on general-media recommendation suggests that personalization may sometimes deamplify sensitive content when it is peripheral to the larger catalog, adult media platforms appear to constitute a structurally different environment in which sexuality, identity, bodily representation, and aggressive performance are central to content organization and monetization. The same recommendation mechanisms that moderate extremes in one domain may therefore intensify them in another.

A second implication concerns platform defaults. If future empirical use of this framework shows that cold-start recommendation outputs already display patterned exposure to gendered, racialized, or aggressive categories, then amplification cannot be explained solely by long-term personalization or user preference learning. Instead, homepage design, default ranking, candidate generation, and interface architecture must be understood as normative forces shaping visibility before any meaningful user history exists.

The framework also highlights metadata as a site of algorithmic power. On adult platforms, categories and tags do not simply describe content; they constitute the semantic space through which recommender systems infer similarity, relevance, and discoverability. When platform taxonomies and user-generated folksonomies converge around highly visible clusters, recommendation outputs may reproduce stereotypes not because the system explicitly intends to do so, but because those associations are already embedded in the input space over which optimization operates.

This has wider implications for how personalization should be understood. Public discourse often treats recommenders as mirrors of user desire, but the framework developed here suggests that recommendation is better seen as a recursive socio-technical process in which the platform classifies content, ranks it selectively, records user interaction with that selective exposure, and then uses those interactions to justify future ranking decisions. In such a loop, the system may not simply detect desire but help stabilize, narrow, and intensify it.

The paper's comparative orientation is also important. By placing adult-platform recommendation in relation to sensitivity-aware general-media benchmarks, the framework creates a way to ask not only whether amplification exists, but whether its structure, magnitude, and mechanism are domain-specific. This prevents the study of adult media from becoming either a niche anomaly or a purely moralized object and instead positions it as a theoretically significant case for understanding exposure, representational harm, and systemic platform bias more broadly.

The discussion also carries a governance dimension. The Digital Services Act requires very large online platforms and search engines to identify, analyze, and assess systemic risks stemming from the design and functioning of their services, including algorithmic systems, and specifically includes risks related to dignity, privacy, non-discrimination, and physical and mental well-being. In adult-platform settings, where recommendation outputs may intersect with intimate behavioral inference and socially sensitive classification, the absence of robust auditing methods creates a gap between legal obligation and empirical assessment. The framework proposed here is intended as one way of narrowing that gap.

At the same time, the paper remains cautious. Black-box audits can reveal structured output patterns, but they cannot by themselves identify the exact internal ranking mechanism, prove platform intent, or eliminate the interpretive challenge involved in deciding which categories should count as aggression, stereotype, or harmful amplification. Adult-platform recommenders should therefore be treated as high-risk and analytically important, but claims about causal architecture must remain proportionate to the evidence actually available.

From an interdisciplinary perspective, this makes adult-platform recommendation a valuable object of study for computer science, platform studies, and philosophy of technology alike. Computationally, it raises the question of how recommendation behaves when sensitive categories are central rather than

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 6| June 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

peripheral to the content domain. Socially and philosophically, it raises the question of whether algorithmic systems merely respond to desire or help organize the normative conditions under which desire becomes legible, amplified, and socially consequential.

7. LIMITATIONS AND ETHICAL PROTOCOL

This paper proposes a methodological framework rather than a completed empirical audit, and several limitations must be stated at both the epistemic and implementation levels. First, the framework is designed for black-box settings and therefore infers platform behavior from observable outputs, metadata structures, and controlled audit conditions rather than from direct access to proprietary ranking models or feature pipelines. Any claim about amplification must therefore be interpreted as an inference from output patterns, not as a reconstruction of internal platform logic.

A second limitation concerns the operationalization of sensitivity itself. In datasets such as ML-DDD and AO3, sensitivity can be anchored in relatively explicit warning taxonomies, but adult-platform environments rarely provide a stable warning ontology. The framework therefore relies on analytically constructed categories derived from titles, tags, platform descriptors, and semantic grouping, which means that the classification of a category as "aggressive," "stereotyped," or otherwise sensitive is partly shaped by researcher judgment. The same issue applies to embedding-based clustering, since cosine-similarity thresholds and grouping choices can encode assumptions about what constitutes conceptual proximity or harm.

A third limitation concerns representativeness. Adult-platform metadata are heterogeneous, unstable over time, linguistically inconsistent, and commercially shaped, while scraping traces and public audit repositories may provide only a partial view of platform behavior. Simulated personas are valuable because they permit controlled comparison, but they do not reproduce the full complexity of real-user histories, affective responses, or long-term social context. The framework should therefore be understood as a structured approximation of amplification dynamics rather than a final ontology of harm.

These limitations make ethics central to the framework. Article 9 of the GDPR states that data concerning a natural person's sex life or sexual orientation are special categories of personal data and that processing such data is in principle prohibited unless a specific legal basis and safeguards apply. Any future empirical implementation of this framework must therefore avoid collecting personally identifiable information, avoid storing raw account-level behavioral traces unless strictly necessary and lawfully justified, and prefer simulated personas, anonymized aggregates, and public metadata over real-user account interception or deanonymization. This is not only a legal issue but also a matter of research integrity given the stigma and privacy risks attached to the domain.

Future implementations should also adopt researcher-protection safeguards. Because auditing adult platforms may expose researchers to repeated encounters with explicit, violent, degrading, or psychologically distressing material, empirical work should rely wherever possible on text-first metadata analysis, headless browsing, blurred or suppressed image capture, and automated logging pipelines rather than repeated human viewing of graphic content. Research teams should further use exposure protocols, role rotation, access controls, and debriefing practices where relevant.

There are also ethical limits to algorithmic auditing itself. Automated auditing can uncover meaningful patterns but may also produce artifacts, incomplete observations, or platform adaptation effects, and it should therefore remain proportionate, minimally invasive, rate-limited, and justified by a clear public-interest purpose. This is consistent with the broader risk-governance logic of the Digital Services Act, which requires systemic-risk assessment to be evidence-based and proportionate to the severity and probability of harm. The ethics section should thus be read not merely as a disclaimer, but as a protocol for any future researcher who seeks to operationalize this framework.

8. ROADMAP FOR EMPIRICAL IMPLEMENTATION

Because the present paper is conceptual and methodological rather than results-driven, its practical credibility depends on whether the proposed framework can be executed in a technically realistic way. This section therefore outlines a two-phase roadmap for empirical implementation: offline representation auditing and live persona-based auditing. Together, these phases convert the framework from a theoretical proposal into a reproducible research blueprint.

8.1 Phase 1: Offline representation auditing

The first phase should be conducted offline using public metadata resources and recommendation modeling that does not require live platform interference. A suitable starting point is the Nikity/Pornhub dataset already identified in the manuscript as a public source of multilingual titles and URLs for lexical, semantic, and categorical analysis. Researchers can construct a metadata table from titles, tags, categories, and inferred semantic clusters, then define a sensitivity-coding layer that maps those metadata features to analytically justified categories such as aggression, racialization, stereotyped gender labeling, or related descriptors.

Once this representation layer is built, researchers can simulate recommendation settings using standard offline recommender architectures such as SVD, ALS, or Neural Collaborative Filtering where interaction structure is available, or content-based ranking where metadata dominate. The goal is not to reproduce a platform's proprietary recommender exactly, but to test whether standard recommendation mechanisms operating over adult-platform metadata mathematically favor some sensitive clusters over others. Using the amplification formalization introduced earlier, researchers can compute baseline prevalence, recommendation prevalence, normalized amplification, and aggregate category amplification to determine whether aggressive or stereotyped semantic groups are overrepresented under otherwise conventional recommendation logic.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 6| June 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

This phase serves several purposes. It enables validation of the category ontology, examination of semantic-grouping choices, calibration of clustering thresholds, and comparison of recommender architectures in a reproducible environment before any live audit is attempted. It also provides a safer methodological space because no live interaction with adult platforms or identifiable real-user data is required.

8.2 Phase 2: Live persona-based auditing

The second phase extends the framework into a live black-box audit of adult-platform recommendation behavior. Here the objective is to observe how the platform's actual interface responds to different behavioral histories, default states, and persona-conditioned interactions rather than to infer patterns from metadata alone. This phase can be implemented with automated headless browsers such as Selenium or Playwright operating under strict rate limits and proportionate audit conditions.

The audit should begin with a small set of predefined personas. At minimum, it should include a clean-browser control persona with no meaningful interaction history and one or more narrow-interest personas that repeatedly interact with a bounded semantic cluster such as an aggressive descriptor family, a stereotyped gender category, or another analytically relevant metadata class. Each persona should operate in an isolated browser environment with separate cookies, storage state, and session variables to prevent contamination across profiles. Scripts should navigate homepages, search surfaces, sidebars, and recommendation carousels at regular intervals and store the resulting recommendation traces in structured text-first form.

A practical proof-of-concept design would observe these personas across a 72-hour window. During that interval, the control persona would record default recommendation outputs without meaningful engagement, while the conditioned personas would follow pre-registered interaction schedules focused on specific semantic clusters. At each observation point, scraped recommendation titles, visible descriptors, and related metadata would be passed through the paper's coding pipeline, including any transformer-based semantic grouping necessary to normalize lexical variation. Amplification scores and persona-differential scores could then be estimated under real platform conditions.

The strength of this second phase is that it can reveal the interaction of default ranking, interface architecture, metadata design, and personalization in a live environment. Its limitation is that, like all black-box audits, it detects structured output differences without conclusively identifying the precise internal mechanism responsible for them. For that reason, the strongest empirical program would combine this live phase with the offline phase rather than treating them as substitutes.

9. CONCLUSION

Adult media platforms should no longer remain peripheral to the study of recommender systems, algorithmic fairness, and platform accountability. They constitute a large-scale, high-risk socio-technical environment in which personalization, intimate behavioral inference, metadata classification, and commercially optimized visibility systems converge around especially sensitive forms of content and identity. For that reason, they require an audit framework distinct from those developed primarily for mainstream media.

This paper has proposed such a framework by combining warning amplification, metadata analysis, semantic grouping, comparative benchmarking, and persona-based black-box auditing into a single methodological design. Its contribution is not to claim completed empirical proof, but to provide a formal, interpretable, and ethically bounded blueprint through which future researchers can test whether adult-platform recommendation systems amplify aggressive, stereotyped, racialized, or otherwise sensitive categories in patterned ways.

This contribution also has regulatory significance. Article 34 of the Digital Services Act requires very large online platforms and very large online search engines to identify, analyze, and assess systemic risks stemming from the design and functioning of their services, including algorithmic systems, and explicitly includes risks involving dignity, privacy, non-discrimination, and physical and mental well-being. The framework proposed here offers one concrete way to operationalize that mandate in a domain where the stakes are high but rigorous auditing methods remain underdeveloped.

The broader implication is that accountability in this domain cannot be achieved through legal theory, ethical critique, or technical metrics alone. Adult-platform recommendation systems sit at the intersection of all three, and they therefore require interdisciplinary methods capable of integrating formal measurement, ethical restraint, and socio-technical interpretation. The framework developed here is intended as a starting point for that agenda: a blueprint to be refined, contested, and empirically implemented by the wider research community in order to make one of the internet's least examined recommendation environments more open to scrutiny and more answerable to norms of public accountability.

REFERENCES

1. *Adler, A. M. (2024). Arousal by Algorithm. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4937597*
2. *Airoldi, M., & Rokka, J. (2022). Algorithmic consumer culture. Consumption Markets & Culture, 25(5), 411–428. https://doi.org/10.1080/10253866.2022.2084726*
3. *Machidon, O. M. (2025). Beyond Algorethics: Addressing the Ethical and Anthropological Challenges of AI Recommender Systems. Journal of Media Ethics, 1–13. https://doi.org/10.1080/23736992.2025.2584435*
4. *Rama, I., Bainotti, L., Gandini, A., Giorgi, G., Semenzin, S., Agosti, C., Corona, G., & Romano, S. (2023). The platformization of gender and sexual identities: An algorithmic analysis of Pornhub. Porn Studies, 10(2), 154–173. https://doi.org/10.1080/23268743.2022.2066566*
5. *Wang, S., & Spronk, R. (2023). "Big data see through you": Sexual identifications in an age of algorithmic recommendation. Big Data & Society, 10(2), 20539517231215358. https://doi.org/10.1177/20539517231215358*
6. *Wang, W., Feng, F., He, X., Wang, X., & Chua, T.-S. (2021). Deconfounded Recommendation for Alleviating Bias*

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 6| June 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Amplification. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 1717–1725. https://doi.org/10.1145/3447548.3467249

7. *Bőthe, B., Tóth-Király, I., Potenza, M. N., Orosz, G., & Demetrovics, Z. (2020). Problematic pornography use: Legal and health policy considerations. Current Addiction Reports, 7(4), 461–469*
8. *European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data , and repealing Directive 95/46/EC . Article 9.ELI: http://data.europa.eu/eli/reg/2016/679/oj*
9. *de Araujo Meirelles Magalhães, F., Calle, I.M. (2026). Consumer Protection in the Digital Age: A Reflection on Regulation 2022/2065 (Digital Services Act – DSA) and Agenda 2030. In: Enes, G., Neves, I., Morais Rocha, T. (eds) A Digital Europe for Citizens. DigEUCit 2023. Springer, Cham. https://doi.org/10.1007/978-3-032-02500-5_10*
10. *Narayanan, A. (2023). Understanding social media recommendation algorithms. Columbia Academic Commons (Columbia University). https://doi.org/10.7916/khdk-m460*
11. *Mansoury, Masoud & Abdollahpouri, Himan & Pechenizkiy, Mykola & Mobasher, Bamshad & Burke, Robin. (2020). Feedback Loop and Bias Amplification in Recommender Systems. 2145-2148. https://doi.org/10.1145/3340531.3412152.*
12. *Meserole, C. (2022, September 21). How do recommender systems work on digital platforms? Brookings. https://www.brookings.edu/articles/how-do-recommender-systems-work-on-digital-platforms-social-media-recommendation-algorithms/*
13. *Rama, I., Bainotti, L., Gandini, A., Giorgi, G., Semenzin, S., Agosti, C., Corona, G., & Romano, S. (2022). The platformization of gender and sexual identities: an algorithmic analysis of Pornhub. Porn Studies, 10(2), 154–173. https://doi.org/10.1080/23268743.2022.2066566*
14. *Tomlein, M., Pecher, B., Simko, J., Srba, I., Moro, R., Stefancova, E., Kompan, M., Hrckova, A., Podrouzek, J., & Bielikova, M. (2022). Black-box Audit of YouTube's Video Recommendation: Investigation of Misinformation Filter Bubble Dynamics (Extended Abstract). Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, 5349–5353. https://doi.org/10.24963/ijcai.2022/749*
15. *Horta Ribeiro, M., Veselovsky, V., & West, R. (2023). The amplification paradox in recommender systems. Proceedings of the International AAAI Conference on Web and Social Media, 17(1), 374–385. https://doi.org/10.1609/icwsm.v17i1.22153*
16. *Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., & Burke, R. (2020). Feedback loop and bias amplification in recommender systems. Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2145–2148. https://doi.org/10.1145/3340531.3412152*
17. *Dataset: https://huggingface.co/datasets/Nikity/Pornhub*