Explainable Creator Intelligence: A Human-Centered Framework for Transparent AI Monetization and Content Recommendation in Social Commerce

Ananya Gill
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
ananyamail@unh.edu

Aamir Mittal
School of Computing, Clemson University, Clemson, SC, USA.
aamirmittal87@clemson.edu

Bavi Srivastava
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
ravisrivastava@uc.edu

Abstract

The rapid convergence of social commerce, artificial intelligence, and digital content creation has given rise to complex socio-technical ecosystems in which creators, platforms, advertisers, and consumers interact under opaque algorithmic regimes. Existing recommendation and monetization systems often prioritize engagement metrics and revenue optimization, leaving creators and users without meaningful explanations for why certain content is promoted or how value is distributed. This paper proposes a human-centered framework termed Explainable Creator Intelligence (ECI), which integrates transparency mechanisms into the core architecture of AI-driven content recommendation and monetization in social commerce. The framework builds upon principles of explainable AI, multi-stakeholder governance, and incentive alignment to design systems that are interpretable, auditable, and fair. We examine structural trade-offs between predictive performance and explainability, and between platform efficiency and creator autonomy. The proposed architecture incorporates content provenance verification, micro-licensing compliance, and path-level safety interventions to ensure robust and accountable operations. We also discuss fairness constraints for recommendation algorithms that mitigate creator bias and echo chambers, and we analyze policy implications for regulatory compliance and creator welfare. Sustainability challenges related to computational overhead and data sovereignty are considered. Through cross-domain comparisons with traditional media and finance, the paper highlights the need for standardized transparency APIs and interoperable governance protocols. The ECI framework offers a blueprint for next-generation social commerce platforms that empower creators while maintaining commercial viability and trust. This work contributes to the growing literature on responsible AI commercialization and provides actionable guidance for platform designers, policymakers, and researchers.

Keywords

explainable AI, social commerce, content recommendation, monetization transparency, creator economy, fairness, governance, multi-stakeholder systems, AI infrastructure.

1. Introduction

Social commerce platforms have transformed the landscape of digital content creation and consumption by enabling direct monetization of user-generated content through recommendations, advertisements, and affiliate mechanisms. Central to this transformation is the deployment of artificial intelligence systems that curate content, match creators with audiences, and allocate revenue based on engagement metrics. However, the opacity of these AI systems raises critical concerns regarding fairness, accountability, and creator autonomy. Creators often lack visibility into why their content is demoted or promoted, how revenue shares are calculated, and what data drives algorithmic decisions. Similarly, consumers cannot evaluate the trustworthiness of recommended content, especially when sponsored or incentivized posts are algorithmically amplified. These issues have led to calls for explainable and transparent AI frameworks that embed human-centered values into the technical infrastructure of social commerce [1], [2].

The concept of explainability in AI systems has been extensively studied in high-stakes domains such as healthcare, finance, and criminal justice [3], where decisions have direct consequences for individuals. In social commerce, the stakes are equally significant, as algorithmic decisions affect creators' livelihoods, consumer trust, and the integrity of digital markets. Yet the unique characteristics of social commerce—rapid content turnover, multi-sided platforms, dynamic incentive structures, and the fusion of social interaction with commercial transaction—demand a specialized framework that goes beyond generic explainability approaches [4]. This paper introduces Explainable Creator Intelligence (ECI), a human-centered framework that integrates transparency into the monetization and recommendation pipeline at multiple levels: data provenance, model interpretability, governance compliance, and user-facing explanations.

The ECI framework is built upon three pillars: interpretable recommendation architectures that provide feature-level and counterfactual explanations; transparent monetization mechanisms that link revenue attribution to specific creator actions and content attributes; and governance protocols that enforce ethical constraints through auditable logs and compliance-by-design principles [5]. We argue that achieving transparency in social commerce requires not only technical solutions but also institutional arrangements that align the incentives of platforms, creators, and consumers. The framework is designed to be modular and scalable, accommodating varying levels of computational resources and regulatory requirements across different jurisdictions.

2. Background and Related Work

The evolution of social commerce has been driven by the convergence of social media, e-commerce, and AI-powered personalization. Early recommendation systems relied on collaborative filtering and content-based methods, which provided some interpretability through explicit rating vectors or feature profiles [6]. However, modern deep learning-based recommenders, including neural collaborative filtering and transformer architectures, achieve higher accuracy at the cost of opacity [7]. In parallel, monetization models have evolved from simple advertising revenue sharing to complex multi-tiered structures involving dynamic pricing, auction mechanisms, and performance-based bonuses.

Explainable AI (XAI) research has produced a variety of methods for post-hoc and ante-hoc interpretability, including LIME, SHAP, attention visualization, and concept bottleneck models [3], [8]. While these techniques have been applied to recommendation systems in e-commerce, they often fail to address the multi-stakeholder nature of social commerce, where creators, platforms, advertisers, and consumers have different explanation needs. For instance, a creator may require a causal explanation for reduced visibility, while a consumer may need a justification for why a sponsored post was recommended. Existing XAI approaches typically focus on single-user or single-objective explanations, neglecting the relational dynamics inherent in creator-platform-consumer triads.

Recent work on fairness in recommendation systems has highlighted biases related to popularity, gender, race, and socioeconomic status [9]. In social commerce, these biases are compounded by the fact that monetization opportunities are often tied to algorithmic exposure, creating feedback loops that amplify existing inequalities. Measures such as demographic parity and equal opportunity have been proposed, but their implementation in dynamic content environments remains challenging. Moreover, fairness without transparency may be insufficient, as stakeholders need to understand the basis for fair treatment in order to trust the system.

The intersection of AI governance and content monetization has been explored in the context of digital rights management, content credentials, and licensing frameworks. The Trusted AI Commercialization Infrastructure for SMBs proposed by Yi (2026) emphasizes a multi-tenant architecture that integrates incentive systems, content governance, and standardized recommendation APIs [10]. This work provides a foundation for building transparent monetization pipelines, particularly for small and medium-sized businesses. Similarly, compliance-by-design approaches using C2PA content credentials and W3C ODRL policies have been suggested to enforce micro-licensing for AI-generated content in social commerce [5]. These contributions highlight the importance of embedding legal and ethical constraints directly into the technical infrastructure.

Safety and robustness are additional dimensions that intersect with explainability. The concept of path-level intervention for large foundation models, as introduced by Shi et al. (2026), offers a method for intervening at computation graph nodes to prevent harmful outputs while preserving model utility [11]. This approach has implications for content recommendation, where interventions may be needed to block objectionable material without degrading the overall user experience. Backdoor defenses in vertical split learning, such as those proposed by Shui et al. (2026), are also relevant for distributed settings where multiple parties collaborate on model training without sharing raw data [12]. These methods contribute to the robustness of AI systems in social commerce.

Finally, software security and vulnerability prioritization frameworks, such as the hybrid SAST-DAST-SCA-IAST approach described by Zhou (2026), provide a model for ensuring that the underlying infrastructure is secure against exploitation [13]. In a context where monetization involves financial transactions and personal data, security is a prerequisite for trust.

3. Theoretical Framework: Explainable Creator Intelligence

The ECI framework is grounded in the recognition that transparency in social commerce is a multi-dimensional construct that encompasses informational, procedural, and outcome-based dimensions. Informational transparency refers to the availability of data and model specifications to stakeholders; procedural transparency concerns the clarity of decision-making processes; and outcome transparency involves the traceability of results to inputs and decisions. The framework operationalizes these dimensions through three interrelated components: explanation generation, incentive transparency, and governance auditability.

Explanation generation in ECI is designed to be stakeholder-adaptive. For creators, explanations focus on the causal factors affecting content visibility, such as audience engagement signals, content novelty, and advertiser demand. Counterfactual explanations—"if you had changed X, your content would have been shown to 20% more users"—provide actionable insights. These explanations are generated using a hybrid approach that combines feature attribution methods with causal models that account for platform-specific dynamics. For consumers, explanations emphasize the rationale behind content recommendation, including relevance signals, sponsorship status, and potential biases. The system employs natural language generation to produce human-readable justifications, which can be tailored to the user's level of technical literacy.

Incentive transparency addresses the monetization side of the framework. Revenue attribution is often a black box in current platforms, with creators receiving aggregated payouts without understanding the contribution of individual content assets or promotional campaigns. ECI proposes a fine-grained ledger that records each impression, click, and conversion event, linking them to specific creator actions and content features. This ledger is maintained using a distributed audit trail, enabling creators to verify revenue calculations. Smart contract-based mechanisms can automate micropayments based on pre-agreed rules, reducing disputes and enhancing trust. The transparency of incentives also extends to advertisers, who can audit the alignment between their spending and actual outcomes.

Governance auditability ensures that the entire system operates within ethical and regulatory bounds. This component builds on compliance-by-design principles, where policies such as age restrictions, content moderation guidelines, and licensing terms are encoded as machine-readable rules and enforced at runtime. The framework incorporates content provenance verification using standards like C2PA, which attach cryptographic signatures to content metadata, enabling consumers and platforms to verify the origin and modification history of digital assets [5]. Additionally, the system maintains an immutable log of all algorithmic decisions, accessible to authorized auditors, to facilitate regulatory review and dispute resolution.

4. System Architecture and Design Principles

The ECI architecture is organized into four layers: the data layer, the inference layer, the explanation and transparency layer, and the governance layer. The data layer handles collection, storage, and preprocessing of content, user behavior, and contextual signals. Privacy-preserving techniques such as differential privacy and federated learning are employed to protect sensitive information, while data lineage tracking ensures that all data transformations are recorded for auditability. The inference layer comprises the core recommendation and monetization models. Instead of a monolithic deep neural network, the architecture adopts an ensemble of interpretable and high-accuracy models, with a gating mechanism that selects the most appropriate model based on the explanation requirements of the user or context. For example, a linear model with feature coefficients may be used for a creator requesting a detailed explanation, while a deep model can be reserved for high-throughput real-time recommendations where interpretability is less critical.

The explanation and transparency layer is the distinguishing feature of ECI. It includes a suite of explanation generation modules that run in parallel with the inference layer, producing both local and global explanations. Local explanations are generated for individual recommendations or monetization decisions, while global explanations summarize overall system behavior, such as the distribution of content exposure across creator groups. This layer also maintains a transparency dashboard accessible to creators and consumers, where they can query past decisions and explore hypothetical scenarios.

The governance layer enforces compliance with ethical and legal standards. It includes a policy engine that interprets rules expressed in formal languages such as ODRL and compares them against content metadata and user attributes. In case of policy violations, the engine can trigger interventions—such as demotion, removal, or human review—at the inference or explanation layer. The governance layer also interfaces with external registries for content credentials and digital rights management.

A key design principle of ECI is modularity: components can be replaced or upgraded independently without disrupting the entire system. This is essential for adapting to evolving regulations and best practices. Another principle is scalability: the explanation and governance layers are designed to operate with minimal latency overhead, using cached explanations and asynchronous auditing. The trade-off between explanation depth and computational cost is managed through adaptive fidelity, where simpler approximations are used for high-throughput scenarios and more detailed explanations are generated on demand.

5. Transparency and Monetization Mechanisms

Monetization in social commerce involves multiple revenue streams: direct sales commissions, advertising revenue shares, subscription fees, and sponsored content agreements. Current practices often obscure the contribution of each stream to a creator's earnings, leading to distrust and dissatisfaction. The ECI framework introduces a transparent monetization ledger that logs every transaction with a unique identifier, timestamp, and attribution vector. Attribution vectors specify the set of content pieces, user interactions, and platform actions that contributed to a revenue event. This granularity allows creators to see, for example, that a particular video generated $50 in ad revenue during the first week, $30 from affiliate links, and $10 from a sponsored integration.

The ledger is maintained using a combination of off-chain databases and on-chain hashes to ensure immutability while preserving scalability. Smart contracts can be used to automatically execute micropayments based on predetermined rules, reducing the need for centralized reconciliation. The transparency of monetization is further enhanced by providing creators with simulated scenarios: what would earnings have been if a different content strategy had been pursued? These simulations rely on counterfactual models trained on historical data, but they are carefully bounded with uncertainty estimates to avoid overconfidence.

Advertisers also benefit from transparent monetization. They can audit the performance of their campaigns at the level of individual impressions and clicks, verifying that they are not being charged for fraudulent or low-quality traffic. The ECI framework incorporates detection mechanisms for click fraud and view bot activity, using anomaly detection and cross-referencing with provenance data. This creates a fairer marketplace where genuine creator efforts are rewarded.

However, transparency in monetization introduces trade-offs. Detailed attribution may reveal proprietary algorithms or sensitive business strategies of the platform. To address this, the framework supports tiered transparency, where aggregated or anonymized views are provided to external auditors, while creators receive personalized but non-revealing information about competitive algorithms. Differential privacy techniques can be applied to the release of aggregate statistics to prevent reverse engineering.

6. Content Recommendation and Fairness

Recommendation algorithms on social commerce platforms influence not only user experience but also the economic opportunities available to creators. Biased recommendations can lead to winner-take-all dynamics, where a small fraction of already popular creators receive disproportionate exposure, while new or niche creators struggle to gain traction. The ECI framework incorporates fairness constraints that aim to balance exposure across creator groups defined by attributes such as size, topic, geographic region, or demographic characteristics. These constraints are operationalized as soft penalties in the recommendation objective function, rather than hard quotas, to preserve recommendation quality.

Fairness in ECI is procedural: the recommendation system must provide explanations for any differential treatment. For instance, if a creator from an underrepresented group receives lower exposure, the system should be able to justify this based on objective factors such as audience taste alignment or engagement projections, rather than opaque algorithmic bias. This requirement encourages the use of interpretable models or hybrid approaches where a transparent model is used for fairness-sensitive decisions.

Another important aspect is mitigating echo chambers and filter bubbles. In social commerce, algorithmic amplification of certain content can polarize consumer preferences and reduce diversity of recommendations. ECI introduces a diversification module that injects serendipity by occasionally surfacing content from outside a user's typical preference neighborhood. The trade-off between personalization and diversity is managed by user-adjustable sliders or system-level defaults, with explanations provided for the selected balance.

Fairness also extends to the monetization of content. Creators should not be penalized for content that is truthful but not aligned with platform profit motives. The framework includes an appeals mechanism where creators can flag perceived unfair treatment, triggering a human-in-the-loop review supported by the transparency dashboard. This mechanism is essential for building trust and ensuring that algorithmic decisions are contestable.

7. Governance, Policy, and Ethical Implications

Deploying an explainable and transparent monetization and recommendation system requires robust governance structures that span technical, legal, and social domains. The ECI framework advocates for a multi-stakeholder governance model where platform representatives, creator councils, consumer advocacy groups, and regulators co-design policies and audit mechanisms. This participatory approach helps ensure that transparency objectives reflect the values of all affected parties.

From a policy perspective, the framework aligns with emerging regulations such as the European Union's Digital Services Act and the AI Act, which require platforms to provide transparency about algorithmic recommendations and content moderation. The compliance-by-design approach of ECI facilitates adherence to these regulations by making policy enforcement an integral part of the system architecture. For example, age-appropriate content recommendations can be enforced through metadata tags and model-level constraints, with auditable logs to demonstrate compliance.

Ethical considerations include the potential for transparency to expose creators to privacy risks or competitive harm. The platform must carefully balance the public's right to know against the creator's right to privacy. ECI addresses this by providing different levels of transparency for different audiences: creators receive detailed personalized reports, while the general public receives aggregated statistics without personally identifying information. Similarly, consumers who opt out of recommendation explanations are not forced to receive them.

Another ethical challenge is the risk of gaming the system. When monetization rules are transparent, bad actors may attempt to exploit them, for example by artificially inflating engagement metrics to gain higher revenue shares. The governance layer includes behavior-based anomaly detection and rate-limiting mechanisms to deter abuse, while still preserving the benefits of transparency for honest participants. The framework also incorporates a feedback loop where governance policies are regularly reviewed and updated based on observed patterns of misuse.

8. Sustainability and Deployment Challenges

Deploying the ECI framework at scale poses significant computational and organizational challenges. The parallel inference and explanation generation layers add overhead to the recommendation pipeline, which may increase latency and energy consumption. Efficient caching strategies, model distillation, and approximate explanation methods can mitigate these costs, but they introduce trade-offs in explanation accuracy. Organizations must decide on acceptable thresholds for latency versus fidelity, and these decisions should be informed by stakeholder input.

Data sovereignty and localization are additional concerns, especially in jurisdictions with strict data protection laws. The ECI framework supports edge deployment and federated learning to keep sensitive data local, but this complicates the maintenance of a unified transparency ledger. Hybrid architectures that combine local explainers with global aggregators can address this issue, though they require careful design to prevent privacy leaks.

Sustainability in the broader sense of long-term viability also depends on the economic model. Transparent monetization may reduce platform profits in the short term if creators demand higher shares or if advertisers reduce spending due to perceived inefficiencies. However, evidence from other industries suggests that transparency can build trust and reduce churn, leading to long-term gains. The framework includes mechanisms for dynamic pricing and revenue-sharing adjustments that adapt to market conditions while remaining transparent.

Another deployment challenge is the cultural shift required within platform organizations. Engineers and product managers accustomed to optimizing for engagement may resist implementing transparency features that could reduce key metrics. Leadership commitment and incentive restructuring are necessary to ensure successful adoption. The ECI framework proposes a gradual rollout, starting with a pilot group of creators and consumers, and using their feedback to refine the system before full deployment.

9. Future Directions

The ECI framework opens several avenues for future research. One promising direction is the integration of large language models for generating explanations that are not only accurate but also empathetic and contextually appropriate. While LLMs can produce fluent text, they also hallucinate, so rigorous validation mechanisms are needed. Another area is the development of standardized transparency APIs that allow third-party auditors to independently verify platform claims. This would require agreement on data formats, access protocols, and verification methods across the industry.

Advancements in causal inference could further enhance the counterfactual explanation capabilities of ECI, enabling more precise attribution of outcomes to actions. Additionally, as social commerce expands into virtual and augmented reality environments, the ECI framework must be extended to handle multimodal content and immersive interactions. Finally, cross-domain comparisons with media, finance, and healthcare can inform best practices for transparency governance in socio-technical systems.

10. Conclusion

Explainable Creator Intelligence represents a necessary evolution in the design of AI systems for social commerce, moving from opaque, profit-driven algorithms to human-centered frameworks that prioritize transparency, fairness, and accountability. By integrating explainability into the core architecture of recommendation and monetization, the ECI framework empowers creators with actionable insights, provides consumers with trustworthy justifications, and enables platform governance that aligns with regulatory and ethical standards. The proposed architecture addresses key structural trade-offs between performance and interpretability, computational cost and transparency depth, and platform efficiency and stakeholder autonomy. While deployment challenges remain, the modular and scalable design of ECI offers a pragmatic path forward. As social commerce continues to grow as a dominant economic force, the adoption of transparent and explainable AI will be essential for sustaining trust, fostering diversity, and ensuring that the benefits of digital content creation are equitably distributed. This work contributes a comprehensive foundation for researchers, practitioners, and policymakers committed to building a fairer and more accountable digital economy.

References

1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138-52160.
2. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.
3. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
4. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).
5. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33-44).
6. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37.
7. Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. ACM Computing Surveys, 52(1), 1-38.
8. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).

9. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1-35.

10. Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.

11. Shi, C., Li, S., Lu, W., Wu, W., Wang, C., Cheng, Z., ... & Chua, T. S. (2026). TraceRouter: Robust Safety for Large Foundation Models via Path-Level Intervention. arXiv preprint arXiv:2601.21900.

12. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59-68).

13. Zhou, D. (2026). AI-Driven Hybrid SAST–DAST–SCA–IAST Framework for Risk-Based Vulnerability Prioritization in Microservice Architectures.

14. Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & van Moorsel, A. (2020). The relationship between trust in AI and trustworthy machine learning technologies. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 272-283).

15. Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudík, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-16).

16. Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. The Information Society, 34(1), 1-14.

17. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).