International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

THE ALGORITHMIC VISIBILITY GAP: A DATA-DRIVEN INVESTIGATION OF REACH INEQUALITY AND ENGAGEMENT DISPARITIES ACROSS SOCIAL MEDIA CREATORS

Ms. Kadiyala Revathi Sri Deepthi

Student, BS-MS(Computer Science), Vishwa Vishwani institute of systems and management
Secunderabad, Telangana, India | Email: deepthikadiyala1507@gmail.com | Mobile:8121164553

ABSTRACT

Social media platforms provide creators with opportunities to publish content and build audiences, but visibility and audience response may be distributed unevenly. This study proposes a data-driven examination of reach inequality and engagement disparities among small, medium, and large creators across multiple platforms. It will use a publicly available dataset, subject to verification of its provenance, coverage, and available variables. Python and Pandas will be used to clean and summarize the data, while descriptive statistics and suitable comparative measures will be applied to examine creator-level reach and engagement patterns. The analysis will distinguish audience size from content-level outcomes and will compare platforms only where definitions and measurement periods are sufficiently compatible. Because the dataset has not yet been selected and analysed, no empirical results or conclusions are reported in this draft. The study is intended to identify whether observed distributions differ across creator-size groups and platforms, and to describe those patterns without treating association as proof that a platform algorithm caused them. Its contribution will depend on the quality and limits of the selected public data. The completed paper should report the dataset, operational definitions, exclusions, statistical procedures, results, and limitations transparently.

Keywords: Social media creators, Algorithmic visibility, Reach inequality, Engagement, Creator size

1. INTRODUCTION

Social media platforms help creators publish content and build audiences, but not all creators receive equal visibility. Feeds, search, recommendations, and sharing shape how audiences encounter content. Researchers have therefore examined visibility as an important part of platform participation (Bucher, 2012; Cotter, 2019).

Creators differ in audience size and in attention received by individual posts. Reach measures exposure, while engagement records responses such as likes, comments, shares, or views. These measures are not interchangeable, and platform definitions and time periods can complicate comparisons.

This study proposes a quantitative comparison of reach and engagement among small, medium, and large creators across platforms, using a public dataset and Python/Pandas. As the dataset has not yet been selected or analysed, results remain pending. Observed differences will be described without assuming that algorithms caused them.

2. REVIEW OF LITERATURE

Bucher (2012) discussed how Facebook's News Feed could shape visibility and create the possibility that content becomes less visible. This work offers a conceptual foundation for examining how platforms affect who is seen.

Cotter (2019) examined Instagram influencers' efforts to interpret platform algorithms, while Bishop (2019) studied YouTube creators' shared beliefs about visibility. These studies illuminate creator experiences but do not prove that particular strategies cause greater reach.

Quantitative research also examines unequal attention. Bärtl (2018) analysed YouTube channel and video views over time, while Lazovich et al. (2023) showed how distributional inequality measures can help describe exposure in recommendation systems.

Wies, Bleier, and Edeling (2023) found a non-linear relationship between follower count and engagement in sponsored Instagram campaigns. Their results are specific to that context. Together, these studies support careful comparisons that distinguish platform metrics and populations.

3. RESEARCH GAP

Prior research considers platform visibility, creator experiences, and unequal attention, often in specific contexts. Further descriptive work can compare reach and engagement by creator size across platforms using a documented public dataset. Such comparisons require compatible measures and cannot reveal proprietary recommendation processes by themselves.

4. OBJECTIVES OF THE STUDY

- To describe reach distribution in the selected dataset.
- To compare reach and engagement across small, medium, and large creators.
- To examine platform differences where measures are comparable.
- To identify data limitations and avoid unsupported causal claims.
- To distinguish descriptive associations from algorithmic causes.

5. RESEARCH METHODOLOGY

Research Design: Quantitative, descriptive comparison using an existing public dataset.

Nature of Data: Secondary data. The dataset source, coverage, collection period, licence, and unit of analysis will be documented after selection.

Study Population and Unit: Creators or creator content included in the dataset; the unit may be an account, post, video, or creator-period record.

Creator Groups: Small, medium, and large groups will use transparent audience-size thresholds, set after reviewing the dataset and reported in the final analysis.

Variables: Only available and clearly defined fields—such as audience size, reach, views, or interactions—will be analysed. Missing or platform-specific measures will be noted.

Tools and Analysis: Python and Pandas will support cleaning and summaries. Descriptive statistics and suitable inequality or group-comparison measures will be selected after checking data quality and distribution.

Comparability and Ethics: Platforms will be compared only when metric definitions and periods align. Necessary data will be used, results reported in aggregate where feasible, and source terms documented.

6. CONCEPTUAL BACKGROUND

Algorithmic Visibility: How platform systems make content discoverable or prominent; this study treats it as a concept unless direct exposure data are available.

Reach: Content exposure, ideally unique accounts reached during a defined period. Views or impressions are proxies and may not represent unique people.

Engagement: Recorded audience actions such as likes, comments, or shares. Counts and rates differ, and any rate must state its denominator.

Reach Inequality: Uneven exposure across creators, described with distributional measures that do not explain its causes.

Creator Size: Audience-size grouping based on a dataset measure such as followers or subscribers; thresholds are study-specific.

7. ANALYSIS AND DISCUSSION

The following subsections describe planned analyses. Results will be added after the dataset is selected and analysed.

7.1 Dataset Profile and Data Quality

The study will report the dataset, platforms, dates, sample, variables, and collection method. Missing data, duplicates, and exclusions will be documented; coverage limits will guide generalization.

7.2 Reach Distribution and Inequality

Reach will be summarized by platform and creator group using suitable distribution statistics. If the data permit, concentration will also be described. Results will remain descriptive unless the design supports stronger inference.

7.3 Engagement Across Creator-Size Groups

Engagement counts and rates will be reported separately. Any rate will identify its denominator, such as reach or follower count. Content format and observation period will be noted where available.

7.4 Comparison Across Platforms

Platforms will be compared only where measures and time periods align; otherwise, they will be analysed separately. Differences cannot by themselves show that algorithms caused unequal reach.

7.5 Interpretation and Limitations

Interpretation will consider sampling, metric definitions, and missing context such as genre, location, or posting frequency. Public metrics alone generally cannot isolate algorithmic effects.

8. MAJOR FINDINGS

Findings are pending dataset selection and analysis. No empirical result is claimed here. The final section should report verified sample details and patterns, and clearly state whether direct reach data were available.

9. SUGGESTIONS / RECOMMENDATIONS

These are preliminary recommendations, not conclusions from completed analysis.

Document dataset sources, inclusion rules, cleaning steps, and metric definitions for reproducibility.

Report reach and engagement separately and define each engagement-rate denominator.

Compare platforms only when metrics and time periods are sufficiently aligned.

Do not describe observed differences as algorithmic causes without evidence that supports causation.

Dataset providers and platforms should clarify metric definitions and coverage limits.

Interpret audience size, reach, and engagement together and in context.

10. CONCLUSION

This study proposes to examine reach inequality and engagement among creators of different sizes across social media platforms. Using a verified public dataset, Python, Pandas, and statistical analysis, it will compare patterns while accounting for differences in measurement. Since the data have not yet been analysed, no empirical conclusion is made. The final interpretation will reflect the dataset's limits and distinguish observed associations from algorithmic causes.

REFERENCES

Bärtl, M. (2018). YouTube channels, uploads and views: A statistical analysis of the past 10 years. Convergence, 24(1), 16–32. https://doi.org/10.1177/1354856517736979

Bishop, S. (2019). Managing visibility on YouTube through algorithmic gossip. New Media & Society, 21(11–12), 2589–2606. https://doi.org/10.1177/1461444819854731

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. New Media & Society, 14(7), 1164–1180. https://doi.org/10.1177/1461444812440159

Cotter, K. (2019). Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. New Media & Society, 21(4), 895–913. https://doi.org/10.1177/1461444818815684

Lazovich, T., Biega, A. J., Ge, Y., Barocas, S., Tardieu, D., & Liao, Q. V. (2023). Measuring disparate outcomes of content recommendation algorithms with distributional inequality metrics. Patterns, 4(1), 100568. https://doi.org/10.1016/j.patter.2022.100568

Wies, S., Bleier, A., & Edeling, A. (2023). Finding Goldilocks influencers: How follower count drives social media engagement. Journal of Marketing, 87(3), 383–405. https://doi.org/10.1177/00222429221125131