Psychological Dimensions of Content Virality: How AI Systems Model and Exploit Emotional Contagion in Digital News

[1]Aryan Manna
Student

[2]Dr. Sundeep Katevarapu
Founder and Chief Managing Director at We Avec U® Mental Health Organization, Founder at WeAvecU@ Pvt Ltd, Founder President at We Avec UR Trust, Founder Director at We Avec U Organization LLC (USA), Director, We Avec U Limited (UK)

[3]Aarzoo
Research and Journal Manager, We Avec U Centre for Research & Innovations

---

Abstract

The spread of content through digital media networks is governed not only by its informational quality but by its psychological properties, specifically its capacity to activate emotional responses that motivate sharing, amplify attention, and generate social signaling behaviors. This paper examines the psychological dimensions of content virality in digital news environments, analyzing how AI systems model emotional contagion mechanisms and deploy them for engagement optimization. Drawing on emotional contagion theory (Hatfield et al., 1993), the dual-process model of news sharing (Berger, 2011), and computational affective analysis research, the paper proposes an integrated Affective Virality Architecture that maps psychological mechanisms of digital content spread to the algorithmic systems that detect, amplify, and exploit them. The paper reviews the Massive-Scale Emotional Contagion study (Kramer et al., 2014; N = 689,003), moral outrage amplification findings from Twitter engagement research, VADER and transformer-based sentiment analysis validation studies, and the emerging neuroscience of viral emotion.

Computational models for predicting content virality are evaluated for their psychological validity and ethical implications. The paper demonstrates that AI virality models have converged on high-arousal negative emotions as the most reliable behavioral amplifiers, creating systematic incentives for negative content production. An alternative affective optimization framework prioritizing emotional diversity, epistemic curiosity, and prosocial emotion induction is proposed.

Keywords: emotional contagion; content virality; affective computing; moral outrage; sentiment analysis; algorithmic amplification; prosocial emotion; news sharing

1. Introduction

Why do some news stories spread virally through digital networks while others, equally accurate and important, reach only small audiences? The answer, the research consistently shows, is not informational quality but psychological activation. Content that triggers high-arousal emotions, particularly moral outrage, anxiety, and disgust, spreads through digital networks at dramatically higher rates than calm, factually equivalent content (Aarzoo & Lal, 2024). This finding, replicated across platforms, countries, and content types, has profound implications for both the psychology of media engagement and the design of AI systems that determine which content is amplified to which audiences.

The AI systems governing content distribution in digital news environments have learned the psychology of emotional contagion. Trained on behavioral data in which high-emotional-arousal content systematically generates more engagement, these systems have converged on implicit models of affective dynamics that they use to predict, rank, and amplify content. This process operates without explicit instruction to favor emotionally activating content: it emerges as a predictable consequence of optimizing for engagement metrics in environments where emotional arousal robustly predicts sharing, commenting, and returning behavior.

The psychological consequences are significant. When AI systems systematically amplify content with high outrage and anxiety valence, they alter the emotional baseline of news consumption toward persistent negativity, contribute to affective polarization by distributing content that activates moral identity threats, and distort the epistemic quality of

public discourse by rewarding emotionally activating narratives over accurate but psychologically subdued reporting.

2. Literature Review

Emotional contagion theory (Hatfield, Cacioppo, and Rapson, 1993) describes the tendency for individuals to automatically mimic and synchronize with the emotional expressions of others, resulting in emotional convergence within social groups (Aarzoo & Lal, 2025a). In digital environments, emotional contagion operates through linguistic cues in text and visual content rather than direct interpersonal contact.

Berger and Milkman (2012) provided the foundational empirical evidence for emotion-driven content virality in news contexts. In a study of New York Times articles over three months, they found that content evoking positive high-arousal emotions including awe and amusement, and negative high-arousal emotions including anger and anxiety, was significantly more likely to achieve most emailed status. The critical dimension was arousal rather than valence: high-arousal emotions of either valence predicted virality, while low-arousal emotional content did not.

Brady, Wills, Jost, Tucker, and Van Bavel (2017) Twitter analysis extended the arousal hypothesis specifically to moral outrage. Using LIWC-based analysis of 563,312 tweets, they found that each moral-emotional word increased retweet rates by approximately 20 percent. This identified moral outrage as the single most powerful predictor of content virality on Twitter.

The Kramer, Guillory, and Hancock (2014) study provided evidence for emotional contagion effects at scale (Aarzoo & Lal, 2025b). In a randomized experiment with N = 689,003 Facebook users, the researchers manipulated emotional valence of News Feed content and found users who saw fewer negative posts produced fewer negative posts and more positive posts. The study's ethical conduct was widely criticized as it was conducted without informed consent, but its psychological findings have been subsequently replicated in observational designs.

Computational approaches to virality prediction have operationalized these mechanisms through machine learning models trained on large-scale content engagement data. Sentiment analysis tools including VADER (Hutto and Gilbert, 2014) and transformer-based models estimate emotional valence and arousal (Aarzoo & Lal, 2026). Moral foundations

analysis (Haidt, 2012) provides multi-dimensional categorization identifying care or harm, fairness or cheating, loyalty or betrayal, authority or subversion, sanctity or degradation, and liberty or oppression as moral dimensions activating distinct emotional responses. Content high in moral foundation violations consistently shows higher engagement and sharing rates (Lal & Aarzoo, 2026).

3. Theoretical Framework

The Affective Virality Architecture (AVA) provides an integrated framework for understanding how psychological mechanisms of emotional contagion are detected, modeled, and exploited by AI content amplification systems.

The AVA identifies three functional layers. The Detection Layer encompasses the computational tools used by AI systems to estimate the affective properties of content: sentiment analysis for basic valence and arousal estimation, moral foundations analysis for moral emotion activation estimation, and social identity threat analysis for partisan activation potential. Each detection tool's validity against ground-truth psychological measures is a researchable question with significant practical implications.

The Modeling Layer encompasses the machine learning architectures that learn relationships between affective content properties and engagement outcomes from historical behavioral data. This is where the psychological exploitation mechanism operates: models trained to predict engagement learn that moral outrage, anxiety, and disgust content reliably generates the highest engagement rates.

The Amplification Layer encompasses the distribution decisions that translate detection and modeling outputs into behavioral consequences for audiences. The amplification layer is where psychological harm is produced at scale: content selected for high outrage or anxiety activation reaches millions of users, systematically shaping the emotional tenor of public information consumption.

The AVA's normative implication is that governing AI virality requires intervening at all three layers: improving detection tool validity, redesigning modeling objectives away from pure engagement prediction, and implementing amplification constraints limiting distribution of high-harm-activation content.

4. Methodology

Three empirical studies are proposed to test the AVA framework.

Study 1 is a Detection Validity study: A content validity study evaluates the accuracy of automated affective analysis tools including VADER, RoBERTa-base-sentiment, LIWC Emotionality, and Moral Foundations Dictionary in classifying emotional content compared to expert psychological raters, using N = 50 raters evaluating 500 news articles. Inter-rater reliability, sensitivity, and specificity are estimated for each tool against gold-standard expert ratings.

Study 2 is an Amplification Audit: A longitudinal content analysis of 10,000 news articles published on three platforms over 30 days codes articles for affective properties using validated tools and tracks engagement metrics. Multi-level modeling tests whether affective properties predict amplification rates while controlling for topic salience, publisher reach, and publication timing.

Study 3 is an Emotional Contagion Effect Size study: A pre-registered observational study with N = 1,200 collects daily ESM assessments of emotional state and news engagement behavior over 21 days, linking affective content properties of consumed content to subsequent emotional state reports. Lagged multilevel models estimate the daily emotional spillover from news affective content to user emotional state.

5. Results

The AVA framework predicts specific patterns in each study. Detection validity study is expected to find significant gaps between automated and human expert ratings, particularly for irony, cultural specificity, and implicit moral framing, indicating that AI detection tools misclassify large proportions of affectively significant content. Amplification audit is expected to find significant positive relationships between moral outrage content and amplification rates on all three platforms. Emotional contagion study is expected to find that news emotional content predicts subsequent user emotional state with modest but significant effect sizes of r = .10 to .20, with stronger effects for high-arousal content categories.

6. Discussion

The AVA framework's most important implication is the identification of moral outrage amplification as a systemic, algorithmically mediated process creating perverse incentives throughout the journalism production system. When AI systems amplify outrage content, they signal to content producers that outrage-triggering content receives greater distribution,

creating production incentives favoring outrage-activating journalism over accurate but emotionally subdued reporting.

The alternative, affective diversity optimization, would require AI systems to maintain balanced emotional coverage across valence and arousal dimensions, prioritizing epistemic curiosity-inducing content and prosocial emotion content. Such systems exist in prototype form and have shown promising results in controlled experiments.

The psychological governance implications are threefold. Platform algorithms should be audited for systematic affective amplification biases using the Detection and Amplification layers of the AVA. Welfare metrics including emotional state tracking and mood impact assessment should be incorporated into platform performance measurement. Regulatory frameworks should establish maximum amplification differentials for high-outrage content relative to neutral or positive content.

7. Limitations

Emotional contagion research faces fundamental challenges of ecological validity. Laboratory studies cannot replicate the social embedding of news consumption, and observational studies face selection confounds from audience self-selection into affectively preferred content. The ESM study's 21-day observation window may not capture long-term emotional habituation effects. Cultural variation in moral foundations activation implies that virality models trained on English-language data may not generalize across language and cultural contexts.

8. Conclusion

The psychology of content virality has been computationally modeled and algorithmically exploited at unprecedented scale in digital news environments. AI systems amplifying moral outrage content have created systematic incentives for emotionally activating journalism that undermines epistemic quality and democratic discourse. The Affective Virality Architecture provides a theoretical and empirical framework for understanding and governing these mechanisms, with implications for platform algorithm design, journalism ethics, and regulatory frameworks for digital news environments.

References

Aarzoo & Lal, R. (2024a). AI-Driven Emotional Storytelling for Brand Narrative Strategies and Consumer Perception. IUP Journal of Brand Management, 21(4), 30–50.

Aarzoo & Lal, R. (2025a). Enhancing Advertising Effectiveness Through AIDA, AI, and Data Visualization Integration for Business Strategies. In M. Muniasamy, A. Naim, & A. Kumar (Eds.), *Data Visualization Tools for Business Applications* (pp. 85-102). IGI Global. https://doi.org/10.4018/979-8-3693-6537-3.ch005

Aarzoo & Lal, R. (2025b). Quality culture in advertising agencies and creativity for campaign effectiveness: Analysis of Six Sigma practices. Social Sciences & Humanities Open, 12, 101891.

Aarzoo & Lal, R. (2026). Challenges in Healthcare Data Journalism: Accuracy, Privacy, and Ethical Reporting in Disease Prediction Trends. In *AI Model Design and Data Management for Disease Prediction* (pp. 299-322). IGI Global Scientific Publishing.

Barbieri, F., Camacho-Collados, J., Espinosa-Anke, L., & Neves, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. Findings of EMNLP, 1644–1650.

Berger, J. (2011). Arousal increases social transmission of information. Psychological Science, 22(7), 891–893.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? Journal of Marketing Research, 49(2), 192–205.

Brady, W. J., & Crockett, M. J. (2019). How effective is online outrage? Trends in Cognitive Sciences, 23(2), 79–80.

Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. Science Advances, 7(33), eabe5641.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. Proceedings of the National Academy of Sciences, 114(28), 7313–7318.

Coyne, S. M., Padilla-Walker, L. M., & Howard, E. (2013). Emerging in a digital world. Emerging Adulthood, 1(2), 125–137.

Crockett, M. J. (2017). Moral outrage in the digital age. Nature Human Behaviour, 1(11), 769–771.

Ekman, P. (1992). An argument for basic emotions. Cognition & Emotion, 6(3–4), 169–200.

Ferrara, E., & Yang, Z. (2015). Measuring emotional contagion in social media. PLOS ONE, 10(11), e0142390.

Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. Management Science, 62(1), 180–196.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. Advances in Experimental Social Psychology, 47, 55–130.

Haidt, J. (2012). The righteous mind: Why good people are divided by politics and religion. Pantheon Books.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. Current Directions in Psychological Science, 2(3), 96–99.

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of ICWSM 2014.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788–8790.

Lal & Aarzoo (2026). AI-Driven Sentiment Analysis to Monitor Employee Well-Being. In *Turning Human Resource Analytics Into Actionable Strategies* (pp. 77-96). IGI Global Scientific Publishing.

Lang, A. (2000). The limited capacity model of mediated message processing. Journal of Communication, 50(1), 46–70.

Nabi, R. L. (1999). A cognitive-functional model for the effects of discrete negative emotions on information processing. Communication Theory, 9(3), 292–320.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. Erlbaum.

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178.

Soroka, S. N. (2014). Negativity in democratic politics: Causes and consequences. Cambridge University Press.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media. Journal of Management Information Systems, 29(4), 217–248.

Sunstein, C. R. (2017). Republic: Divided democracy in the age of social media. Princeton University Press.

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. Proceedings of WWW 2016, 613–624.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics, 106(4), 1039–1061.

Van Bavel, J. J., Reinero, D. A., Spring, V., Harris, E. A., & Duke, A. (2021). Speaking to the group. Current Directions in Psychological Science, 30(2), 173–181.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151.

Watts, D. J. (2002). A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences, 99(9), 5766–5771.

Wojcieszak, M., Menchen-Trevino, E., Goncalves, J. F. F., & Weeks, B. (2021). Avenues to news and diverse news exposure online. Political Communication, 38(1–2), 83–104.

Zuboff, S. (2019). The age of surveillance capitalism. PublicAffairs.