THE AMPLIFYING MIRROR: PERSONALITY AND SOCIAL MEDIA PLATFORMS

Leif Pesola

Department of Psychology, Eastern Illinois University, USA

ABSTRACT

Social media platforms function as distributed systems that amplify preexisting personality dispositions through algorithmic filtering and feedback. Rather than creating identity, these systems intensify patterns by shaping exposure and reinforcing what users repeat. Evidence from personality psychology and work on algorithmic personalization suggests that stable traits predict how individuals interpret and respond to platform cues. Over time, repeated interaction with curated feeds and reward signals can influence performance orientation and perceived self-consistency. This narrative review synthesizes psychological and computational findings to describe algorithmic architectures as an amplifying mirror that strengthens dispositional tendencies through ongoing feedback within large-scale, algorithmically mediated environments.

KEYWORDS

Personality Traits, Algorithmic Personalization, Distributed Platforms, Digital Identity, Feedback Loops

1. INTRODUCTION

Social media rarely invents a new self. It turns up the volume on what is already there and hands it a brighter stage. Algorithms personalize exposure to reflect traits that existed long before sign-ups and notifications, as Lee et al. [1] observed in their work on algorithmic personalization and self-concept. This paper suggests that dispositional traits, especially extraversion, neuroticism, conscientiousness, and darker profiles, reliably relate to how individuals engage with and are shaped by social media. Algorithms perform much of the selection work, sorting what we see and what we miss. Research on filtering and clustering shows that repeated exposure can narrow what people attend to and strengthen what they already believe [2], [3].

Expression becomes a self-reinforcing experiment. This system of feedback through likes, comments, reposts, and reactions becomes a digital mirror that trains people to see themselves through shifting signals of approval, shaping both self-esteem and identity as those signals rise and fall [4], [5], [6].

To clarify terms used throughout, “mirror with gain” describes a feedback loop that amplifies what already exists. The platform reflects the self back, but with added volume. Traits that draw engagement grow sharper and more visible, while quieter parts of the person fade into the background. “Performance economy” names the incentive system built around that loop, where visibility itself becomes value. Likes, shares, and comments can turn self-presentation into a kind of currency, shaping expression less by conviction and more by what provokes response. In this review, the amplifying mirror is used as a conceptual integration rather than a mechanistic claim, intended to describe how algorithmic feedback structures interact with stable personality traits to intensify existing patterns of expression rather than to specify causal pathways.

Every post and reaction, every scroll, can contribute to how the mind defines itself in the moment, keeping self-concept and self-esteem in motion rather than fixed [4], [7]. The design narrative emphasizes connection, but evidence suggests the lived experience often drifts toward pressure and strain, especially for younger users who carry fragile identity work into feeds that never sleep [8].

This loop resists clean division because human attention and algorithmic curation keep shaping one another in real time. These dynamics create an environment in which visibility becomes the organizing force. Self-presentation starts drifting into a role, shaped by incentives that favor patterns the system already knows how to reward [2], [9]. Theoretically, this helps frame how platforms push presentation to the foreground.

This paper is structured as a narrative review integrating psychological research on personality traits with research on algorithmic filtering and large-scale platform design. Sources were drawn primarily from peer-reviewed literature published between 2021 and 2025, with earlier foundational work included when it provides necessary theoretical grounding. Searches were conducted using combinations of terms such as personality traits, social media use, recommendation systems, algorithmic personalization, feedback loops, and digital identity. Studies were prioritized when they reported measurable associations between traits and online behavior, examined platform-level reward or ranking dynamics, or offered mechanism-level accounts relevant to how exposure is shaped at scale. Because this topic spans psychology, computer science, and human-computer interaction, the goal is conceptual integration rather than meta-analytic synthesis.

2. LITERATURE REVIEW

2.1. Relevance to Distributed and Parallel Systems

Social media platforms can be understood as large-scale distributed systems that process continuous streams of user behavior within predictive feedback architecturesin real time. Recommendation pipelines are implemented on distributed, parallel infrastructures that algorithmically curate content at massive throughput, shaping what becomes visible to individual users. As Lee et al. [1] describe, algorithmic personalization forms a dynamic predictive environment in which identity expression and system-level computation evolve together. The amplification of personality traits emerges from distributed algorithms that impose relevance ordering and behavioral reinforcement at scale[2], [9]. By examining identity formation within these computational structures, this review aligns with research on how large-scale digital infrastructures influence human behavior.

In practice, personalization occurs through large-scale ranking and recommendation pipelines that rely on collaborative filtering alongside engagement prediction, supported by feedback-driven optimization [1], [2]. These systems form reinforcement structures through repeated updates, shaping what users see and what they return to over time [2], [10]. Applied discussions of excessive use and attentional capture describe how these feedback dynamics can strengthen habitual engagement patterns and narrow attention [11], [12].

2.2. Personality Traits and Platform Engagement

This review draws from empirical and theoretical work on personality and algorithmic curation, focusing on how digital identity forms within major platforms. Recent research from roughly 2021 to 2025 anchors the review, with earlier foundational work such as Digman [13] providing

the theoretical spine. Included studies showed measurable associations between traits and online behavior, often with implications for mental health. Where possible, peer-reviewed, psychology-focused journals and meta-analytic or multi-study designs are emphasized. Perspective and systems papers are used for mechanism context rather than causal claims.

Once the system begins to sort the feed, it starts feeding back certain behaviors more often. The patterns it likes grow louder. The ones it ignores thin out. Lee et al. [1] describe how algorithmic ranking influences both the material shown to users and how users become positioned within the system's predictions. This may narrow attention and distort identity, tightening the fit between content and self-concept while polishing a public mask.

Joseph [14] notes that personalized relevance can feel like attention or concern, even though it is generated mechanically. This can result in rising comparison and a curated persona that drifts from the private self, especially when identity is actively negotiated through online self-presentation [15], [7]. User actions and platform adjustments tend to shape each other over time, as noted by Ognibene et al. [10]. Lee et al. [1] call this an algorithmic crystal, a system that refracts the self through its own predictions. Metzler and García [2] observe that user tendencies and algorithmic patterns often shift together in ways that shape what becomes visible.

Across studies, personality stays steady. The same temperaments we see offline show up online. Large-sample work suggests that higher levels of neuroticism or extraversion are often tied to more intense and sometimes more problematic patterns of social-media involvement [16], [17]. Meta-analytic work suggests that conscientious and agreeable users usually post less and report fewer issues with compulsive use, pointing to a more measured engagement [18], [19]. Extraversion consistently predicts more visible and interactive behavior online, especially in environments designed for public-facing interaction [18], [20]. Introverted users participate differently, often maintaining smaller circles or remaining quiet observers even when actively present [8].

2.3. Feedback Dynamics and Identity Expression

Even basic signals such as likesappear to recruit reward-related circuitry and shape behavior over time, a pattern visible in both behavioral and neural findings [5], [6]. Brady et al. [21] showed that posts using moral-emotional language spread more widely, which supports the idea that even small bursts of feedback can reinforce checking behavior. Receiving approval activates reward-related parts of the brain, a finding highlighted by Izuma et al. [22]. Although the studies differed in their methods, the underlying pattern was surprisingly consistent.

Traits linked to narcissism, Machiavellianism, or psychopathy can show up online as manipulative or antagonistic behavior, as the research indicates [23], [24]. When sadism joins this triad, it completes the machinery. Dark Tetrad tendencies have been associated with trolling behaviors, which Buckels et al. [23] describe as a calculated cruelty. Pleasure replaces empathy. The suffering of others is the satisfaction.

Festinger [25] proposed that people evaluate themselves by comparing with others, a process that translates directly to online spaces where every post invites comparison. Goffman [26] framed social interaction as performance, suggesting that people change their behavior when they feel observed compared to when they are not observed. Paliszkiewicz and Mądra-Sawicka [27] note that professional platforms often amplify impression management, rewarding polish over sincerity. Additionally, people sometimes react negatively to overt moral display, a point raised by Minson and Monin [28], creating suspicion rather than admiration. Brady et al. [21] demonstrated that posts containing moral and emotional language were shared more widely

across networks, encouraging performance over quiet practice. Lim and Tan [29] frame this as an evolutionary mismatch, where an old hunger for belonging is now tied to a machine that never sleeps. Together, these findings point to how algorithms and reward patterns lean on a person's temperament and slowly shape the way they become visible online.

2.4. Algorithmic Architecture of Feedback Loops

The amplifying mirror described throughout this review can be understood as an emergent property of adaptive algorithmic infrastructures that continuously update in response to user behavior. Ranking models and predictive systems refine exposure through distributed computation, adjusting relevance signals in real time [1], [2]. Patil et al. [9] highlight how algorithmic amplifiers can escalate or suppress trends depending on feedback. Ognibene et al. [10] demonstrate how recommendation pipelines integrate behavioral signals to tune future predictions. Together, these mechanisms form a parallel feedback structure in which repeated exposure reinforces traits the system learns to reward.

3. ANALYSIS AND DISCUSSION

3.1. Trait Expression and Reinforcement

Personality sits at the center of the aforementioned loop. It governs what people seek and return to, including what they quietly tolerate. Behavioral differences online trace back to stable dispositional traits. Gahlot and Imran [15] describe extraversion as the most visible amplifier, yet the deeper pattern is psychological hunger. Even a small pulse of approval can feel like connection, which is why the pull back to the feed often shows up before someone notices it. Over time, brief cues of approval can take on the emotional weight of genuine connection, a pattern that shows itself in reward studies like Sherman et al. [6]. When those signals fade, users often describe a kind of unease because the feedback is no longer mirroring connection but standing in for it. As a narrative review, these mechanisms function as integrative interpretations rather than causal claims.

Introverted users engage with platforms differently. For many, presence happens through observation rather than performance. The platform seems to follow the person's baseline orientation rather than pushing it in a new direction, which lines up with earlier work such as Correa et al. [20]. Conscientious and agreeable users tend to move against that current. These steadier temperaments might introduce a quieter rhythm. They do not erase it, but they seem to dull its pull, almost as if internal structure moderates external pressure. In short, frequent use does not make a person extroverted. Rather, platforms reinforce outward-seeking tendencies already present, while offering introverted users a narrower path of engagement. Once that pattern settles in, it seems to open the door for darker profiles to show up more clearly.

3.2. Dark Profiles and Moral Signaling

The expression of darker personality profiles represents one pathway through which algorithmic reinforcement operates, not a dominant or universal outcome of platform use. Traits associated with antagonism and manipulation have been linked to trolling and exploitative interaction styles in online environments [23], [24]. What emerges is a stimulus loop that sustains attention while loosening accountability, especially when identity is partially obscured and social costs are delayed. Platform metrics optimize for engagement rather than prosocial value, which means harmful or polarizing content can be elevated even when attention reflects disapproval rather than endorsement [2], [9], [21]. In some cases, behaviors once constrained by proximity gain reach

through visibility, and adaptation accelerates as older social constraints fall away. As an interpretive lens rather than empirical evidence, Jung's concept of the shadow offers a way to think about how disinhibited expression becomes more visible under reduced social constraint [30]. These patterns should be understood as selective amplification and increased visibility of antagonistic tendencies rather than as evidence of rising trait prevalence, reflecting incentive alignment within platform environments rather than dispositional change.These dynamics suggest that performance can become habitual and breadth can crowd out depth, with outcomes shaped by both platform design and individual context [2], [10].

3.3. Identity Consequences

This amplification loop does not affect everyone equally. Its strength depends on the context of both platform and user. Smaller, more personal networks may lessen performance pressure because reputation stays linked to real relationships. Strong anchors outside the screen, including family ties or steady community roles, can keep digital feedback in proportion.

Design choices matter as well. Chronological or randomized feeds can widen exposure and may weaken the pull of reward. Hidden metrics can reduce comparison, while added friction such as delayed feedback slows impulsive performance. Moderation also shapes the tone. When expectations are clear and enforced, attention drifts less toward spectacle. These structural elements might clarify why identity outcomes vary so widely. The platform's shape often predicts the psychological shape that follows.

When a platform reduces the visibility of performance metrics and slows immediate feedback, trait expression remains, but the reinforcement behind it weakens. As reinforcement fades, the demand to keep performing loses fuel. Once that happens, earlier psychological patterns tend to show themselves more plainly, a point that fits the direction of work by Metzler and García [2].
Users often report indifference to attention cues, yet behavioral patterns suggest that even small feedback signals can reinforce return loops. Evidence from neural and behavioral research indicates that social feedback can recruit reward-related processing and shape engagement over time, which helps explain why checking behavior can become routinized [6], [22]. Review evidence also suggests that repeated feedback exposure can influence how people interpret social approval cues [5]. As the loop stabilizes, self-presentation can drift toward what reliably produces feedback, which may increase impression management and reduce perceived authenticity for some users [15], [7], [27]. Similar dynamics can appear on professional platforms where reputational incentives and visibility cues shape posting norms, even when the stated goal is career signaling rather than social belonging [27]. When moral-emotional language is widely shared, it can also become amplified by engagement incentives, which may shift interpretation toward motive and intensify skepticism toward overt display [21], [28].

Tension often pushes users to question motive, especially when moral language functions as a kind of social currency, a direction consistent with reactions documented by Minson and Monin [28]. There is a growing discussion of discomfort with moral signaling and display. Moral display turns fragile when the audience becomes the point rather than the witness. When people post virtue, it rarely lands as virtue. It lands as proof. Goodness becomes a form of marketing. Most do not intend it that way, but digital culture rewards visibility rather than integrity. What begins as a moment of gratitude or conviction can easily become a bid for validation.

4. CONCLUSIONS

Joseph [14] found that social media platforms act as extensions of personality. They reinforce pre-existing dispositions through repeated filtering. Their systems reward surface engagement more than sincere connection. Metzler and García [2] show how social tendencies and algorithmic structure work together to shape attention patterns. Rodilosso [31] adds that systems built for engagement often elevate more attention-grabbing material.

The following synthesis reflects an interpretive integration of findings rather than a normative claim about individual cognition.Alter et al. [32] showed that when information feels easy to process, people rely more on intuition and are less likely to engage in slow, deliberate reasoning. Reducing cognitive effort becomes reflex. Independent appraisal weakens. Fluency masquerades as truth. Once cognition tilts, identity becomes easier to steer from the outside. Pariser [33] argued that personalized filtering reduces exposure to different points of view, which can tighten a person's informational environment over time. Content confirming prior views rises. Opposing material sinks. Culture appears uniform even as each feed insists on uniqueness. Taken together, work on echo chambers and engagement suggests that platform metrics tend to favor what users already recognize: content that fits prior beliefs moves more easily through networks, while unfamiliar or more reflective material often travels more slowly [31], [34], [35]. Conceptually, this fits Alter et al.'s argument [32] that cognitive ease can create a feeling of understanding without actually deepening it.

This review frames algorithmic personalization as a feedback system that can amplify stable dispositional tendencies by shaping exposure and reinforcing repeated behaviors [1], [2], [9]. Because the evidence base spans correlational trait research and mechanism-oriented platform studies, causal claims should be treated cautiously and evaluated alongside context such as moderation strength and network structure [2], [10]. Work on filtering and echo chambers suggests that personalization can reduce exposure diversity over time, which may strengthen prior beliefs and simplify appraisal in ways that feel subjectively coherent even when understanding has not deepened [33], [34], [35]. Future research should connect personality-linked differences in engagement to identifiable properties of ranking systems, including exposure diversity and feedback timing, using designs that separate selection effects from amplification effects [1], [2], [31]. Clarifying these mechanisms would strengthen theory and support platform designs that reduce harmful reinforcement while preserving legitimate self-expression [2], [12].

ACKNOWLEDGEMENTS

The author would like to thank the faculty of the Department of Psychology at Eastern Illinois University for their support and guidance.

REFERENCES

[1] A. Y. Lee, H. Mieczkowski, N. B. Ellison, and J. T. Hancock, "The algorithmic crystal: Conceptualizing the self through algorithmic personalization on TikTok," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, pp. 1-26, 2022.

[2] H. Metzler and D. García, "Social drivers and algorithmic mechanisms on digital media," *Perspect. Psychol. Sci.*, vol. 19, no. 5, pp. 735-748, 2024.

[3] X. Song, S. Guo, and Y. Gao, "Personality traits and their influence on echo-chamber formation in social media," *Front. Psychol.*, vol. 15, Art. no. 1323117, 2024.

[4] Y.-H. Chen, "State self-esteem responses to social media feedback loops," *Front. Psychol.*, vol. 16, Art. no. 1625771, 2025.

[5] A. R. Dores et al., “Effects of social feedback through the ‘Like’ feature on brain activity,” *Healthcare*, vol. 13, no. 1, Art. no. 89, 2025.
[6] L. E. Sherman et al., “The power of the like in adolescence,” *Psychol. Sci.*, vol. 27, no. 7, pp. 1027-1035, 2016.
[7] V. Pérez-Torres, “Social media as a digital social mirror,” *Curr. Psychol.*, vol. 43, no. 26, pp. 22170-22180, 2024.
[8] S. Arora et al., “Psychological impacts of algorithmic social media on teenagers,” in *Proc. IEEE Digital Platforms and Societal Harms (DPSH)*, Washington, DC, USA, 2024, pp. 1-7.
[9] S. Patil et al., “Trend amplification or suppression,” *J. Cyber Policy Ethics*, vol. 2, no. 1, pp. 1-18, 2024.
[10] D. Ognibene et al., “Well-being-aware recommendation algorithms,” *Front. Artif. Intell.*, vol. 5, Art. no. 654930, 2023.
[11] A. O. Mujica et al., “Addiction by design,” *Med. Res. Arch.*, vol. 10, no. 2, pp. 1-29, 2022.
[12] C. Voinea et al., “Digital slot machines,” *Topoi*, vol. 43, no. 3, pp. 685-695, 2024.
[13] J. M. Digman, “Higher-order factors of the Big Five,” *J. Pers. Soc. Psychol.*, vol. 73, no. 6, pp. 1246-1256, 1997.
[14] J. Joseph, “The algorithmic self,” *Front. Psychol.*, vol. 16, Art. no. 1645795, 2025.
[15] V. Gahlot and M. A. Imran, “Personality and online persona curation,” *Int. J. Appl. Res.*, vol. 11, no. 5, pp. 350-356, 2025.
[16] E. Ahmed and S. Ahmed, “Social media addiction and personality,” *Curr. Opin. Psychiatry*, vol. 38, no. 1, pp. 72-77, 2025.
[17] D. Marengo et al., “Big Five traits and social media addiction,” *Addict. Behav.*, vol. 102, Art. no. 106110, 2020.
[18] H. Lin et al., “Big Five and information sharing,” *PLOS ONE*, vol. 19, no. 6, Art. no. e0303770, 2024.
[19] A. M. Weinstein, “Problematic social networking,” *Front. Psychiatry*, vol. 13, Art. no. 1106004, 2023.
[20] T. Correa et al., “Personality and social media use,” *Comput. Human Behav.*, vol. 26, no. 2, pp. 247-253, 2010.
[21] W. J. Brady et al., “Emotion shapes diffusion of moralized content,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 28, pp. 7313-7318, 2017.
[22] K. Izuma et al., “Social and monetary reward processing,” *Neuron*, vol. 58, no. 2, pp. 284-294, 2008.
[23] E. E. Buckels et al., “Trolls just want to have fun,” *Pers. Individ. Differ.*, vol. 67, pp. 97-102, 2014.
[24] D. L. Paulhus and K. M. Williams, “The Dark Triad of personality,” *J. Res. Pers.*, vol. 36, no. 6, pp. 556-563, 2002.
[25] L. Festinger, “A theory of social comparison,” *Human Relations*, vol. 7, no. 2, pp. 117-140, 1954.
[26] E. Goffman, *The Presentation of Self in Everyday Life*. New York, NY, USA: Anchor Books, 1959.
[27] J. Paliszkiewicz and M. Mądra-Sawicka, “Impression management on LinkedIn,” *Management*, vol. 11, no. 3, pp. 203-212, 2016.
[28] J. A. Minson and B. Monin, “Do-gooder derogation,” *J. Exp. Soc. Psychol.*, vol. 48, no. 1, pp. 200-207, 2012.
[29] A. J. Lim and E. Tan, “Evolutionary mismatches and social media,” *Evol. Psychol. Sci.*, vol. 10, no. 3, pp. 212-223, 2024.
[30] [30] C. G. Jung, *Aion*. Princeton, NJ, USA: Princeton Univ. Press, 1959.
[31] E. Rodilosso, “Filter bubbles and polarization,” *Philos. Technol.*, vol. 37, Art. no. 71, 2024.
[32] A. L. Alter et al., “Overcoming intuition,” *J. Exp. Psychol. Gen.*, vol. 136, no. 4, pp. 569-576, 2007.
[33] E. Pariser, *The Filter Bubble*. New York, NY, USA: Penguin Press, 2011.
[34] M. Cinelli et al., “The echo chamber effect,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 118, no. 9, Art. no. e2023301118, 2021.
[35] [35] M. Mosleh et al., “Cognitive reflection and Twitter behavior,” *Nat. Commun.*, vol. 12, Art. no. 921, 2021.

AUTHORS

Leif Pesola is an undergraduate in psychology at Eastern Illinois University. His writing engages personality theory and research on algorithmic personalization, with particular attention to how digital systems shape self-presentation and self-understanding.