

“BLACK MIRROR” AND THE DIGITAL “SELF”: A PSYCHOLOGICAL PERSPECTIVE ON DIGITAL IDENTITY

Beka Dadeshkeliani

Affiliated Assistant-Professor at International Black Sea University,
2 David Aghmashenebeli Alley, 13th km, 0131
Tbilisi, Georgia

Abstract.

Digital identity is an ongoing psychosocial process, not merely a representation of an offline person, through which individuals present, evaluate, and revise the self. This critical integrative review examines self-presentation, authenticity, social comparison, reflected appraisal, and online–offline self-discrepancy. It synthesizes peer-reviewed research published primarily between 2010 and July 2026 while retaining foundational theoretical works. Goffman’s dramaturgical model, Cooley’s looking-glass self, Higgins’s self-discrepancy theory, Foucault’s technologies of the self, Walther’s hyperpersonal model, and Turkle’s networked identity are integrated into a multilevel framework. The evidence indicates that social media effects depend on platform affordances, audience structure, user motives, developmental stage, and alignment between self-concept and expression. Authentic expression and supportive interaction may enhance well-being, whereas idealization, upward comparison, approval dependence, and persistent audience monitoring may contribute to psychological distress. The “black mirror” is developed as a non-fictional metaphor for the screen as stage, social mirror, archive, and algorithmic environment. Digital identity is therefore conceptualized as a negotiated relational system rather than a simple opposition between “real” and “virtual” selves. Future research should prioritize longitudinal, cross-cultural, platform-sensitive, and person-specific designs.

Keywords: *authenticity; digital identity; online self-presentation; social comparison; social media*

Introduction

Digital environments have become ordinary settings for identity work. Global estimates counted 5.66 billion active social media user identities in October 2025, while Georgia had approximately 3.05 million such identities, equivalent to 80.1% of the national population; these estimates refer to accounts or user identities and should not be interpreted as unique persons [1, 2]. The scale of participation matters because profiles, photographs, messages, reactions, search histories, and platform-generated recommendations now form part of the social conditions under which people understand themselves and are understood by others. Digital identity is therefore not merely a technical record or a deliberate profile. It is a psychologically consequential configuration of self-descriptions, behavioral traces, audience responses, and platform classifications.

The phrase *black mirror* is used here as an original analytical metaphor rather than as a reference to, or analysis of, a television series. When a screen becomes dark, it returns a faint image of its viewer; when active, it returns a socially and algorithmically mediated image

composed of posts, metrics, memories, recommendations, and other people's reactions. The metaphor captures a central paradox. Digital media increase the individual's capacity to edit and distribute a self-representation, but they also expose that representation to persistent evaluation and to systems that sort visibility. The screen operates simultaneously as a stage on which identity is performed, a mirror through which identity is evaluated, an archive that preserves earlier selves, and an environment that structures what can be seen and rewarded.

Early research on online interaction sometimes treated the digital domain as separate from embodied social life. That distinction is increasingly difficult to sustain. Facebook research showed that online identity could be anchored in existing relationships rather than detached from them [3], while research on social media exhibitions demonstrated that profiles are persistent collections of content that can be viewed by audiences beyond the moment of posting [4]. Recent social identity theorizing likewise treats digital identities as products of intersecting online and offline group memberships, while recognizing that platform features introduce conditions not captured fully by theories of offline identity [5]. Contemporary identity consequently moves across contexts: a professional profile, a private messaging group, an image-based feed, and an anonymous forum may foreground different facets of the same person. These differences need not imply deception. They may reflect normal role differentiation, audience adaptation, or experimentation. Psychological difficulty becomes more likely when the management of these facets is experienced as compulsory, incoherent, or dependent on continuous approval.

The literature does not support a simple claim that social media is either harmful or beneficial. Reviews consistently find heterogeneous and often small average associations between use and well-being, partly because total "screen time" combines different activities, motives, relationships, and vulnerabilities [6, 7, 8]. Active conversation with supportive others is psychologically different from passive exposure to idealized images, just as voluntary self-expression differs from anxious monitoring of likes. A useful account must therefore move from the amount of use to the mechanisms and contexts of use.

This review addresses four research questions:

- [1] What psychological mechanisms organize digital identity in contemporary online environments?
- [2] How do social media affordances shape authenticity, self-presentation, and self-evaluation?
- [3] Which psychological and social theories best explain the relationship between digital identity and well-being?
- [4] How can the "black mirror" metaphor integrate individual, relational, and platform-level dimensions of digital identity?

The article aims to make two contributions. First, it critically synthesizes research that is often separated into literatures on impression management, social comparison, authenticity, adolescent mental health, and mediated communication. Second, it proposes a multilevel model in which digital identity emerges from the interaction of self-processes, audiences, and platform affordances. This approach avoids both technological determinism, which assigns causal power to platforms alone, and an exclusively individual account that overlooks how design features structure attention and social reward.

Method

Review Design

The study used a critical integrative review design. This form of review is appropriate when the goal is to compare theories, interpret heterogeneous empirical evidence, and develop a conceptual synthesis rather than calculate a pooled effect size. The analysis combined thematic organization with theoretical critique. It was not a systematic review or meta-analysis, and no claims of exhaustive coverage or formal risk-of-bias scoring are made.

Search Scope and Eligibility

The primary publication window was January 2010 through July 2026, covering the consolidation of social networking sites and the subsequent growth of image-based, mobile, and algorithmically curated platforms. Earlier works were retained only when they provided foundational concepts that contemporary studies explicitly develop: the looking-glass self, dramaturgical self-presentation, social comparison, self-discrepancy, technologies of the self, computer-mediated communication, and early work on online identity.

Search concepts combined terms relating to identity and self-processes with digital settings: “digital identity,” “online identity,” “social media self-presentation,” “online authenticity,” “context collapse,” “social comparison,” “self-esteem,” “self-discrepancy,” “digital stress,” and “adolescent mental health.” Scholarly discovery was oriented toward literature indexed or commonly discoverable through Scopus, Web of Science, PsycINFO, Google Scholar, and publisher databases. Reference chaining from influential reviews and empirical studies was used to identify foundational and closely related work. Bibliographic metadata and DOI strings were checked against publisher or DOI-registration records during manuscript revision.

Sources were eligible when they (a) examined identity, self-presentation, self-evaluation, authenticity, or related psychological outcomes in digitally mediated settings; (b) presented empirical evidence, a review, or a theory relevant to the research questions; and (c) were peer-reviewed articles, scholarly books, or chapters from academic publishers. Two DataReportal reports were retained solely to describe global and Georgian participation; they were not treated as psychological evidence. Material centered on fictional episodes was excluded because the “black mirror” functions in this article as a conceptual metaphor rather than an object of textual analysis.

Analytical Procedure

Evidence was coded into four thematic domains: (a) performance and audience management; (b) reflected appraisal, social comparison, and self-evaluation; (c) authenticity, self-discrepancy, and psychological strain; and (d) platform affordances, development, and culture. Within each domain, the review distinguished descriptive findings from causal claims. Cross-sectional associations were not interpreted as evidence that social media behavior produced an outcome. Greater weight was given to longitudinal, experimental, and review-level evidence, while qualitative and theoretical studies were used to clarify mechanisms and meaning.

The final synthesis was guided by three questions: What is the unit of analysis the person, interaction, audience, or platform? What mechanism connects digital practice to identity or

well-being? What alternative explanation remains plausible? This procedure helped separate robust conclusions, such as the importance of social comparison and feedback-seeking, from broader cultural assumptions that are not yet supported by adequate longitudinal or cross-cultural evidence.

Methodological Limitations

The review has four limitations. First, it is integrative rather than systematic, so the corpus cannot be assumed to include every relevant publication. Second, English-language scholarship and Western platforms are overrepresented. Third, the rapid evolution of platforms complicates comparison across studies conducted under different interfaces and norms. Fourth, constructs such as authenticity, active use, and self-presentation are operationalized inconsistently. These limitations make conceptual precision and cautious causal language essential.

Theoretical Framework

Self-Presentation as Performance and Exhibition

Goffman's [9] dramaturgical approach conceptualizes social interaction as impression management before an audience. Individuals assemble a "front" suited to a setting while preserving regions in which the public performance can be suspended. Applied to social media, the framework explains why profile photographs, biographies, captions, and visible affiliations are selected with an audience in mind. Yet online identity cannot be reduced to a live theatrical performance. Hogan [4] distinguishes performance from exhibition: a post may be created in one moment but persist as an artifact that is later accessed, recirculated, or recombined. This temporal persistence expands the audience and makes identity management partly archival.

Context collapse intensifies the problem. Audiences that are normally separated family members, colleagues, friends, students, and strangers can converge around the same content. Users respond by imagining an audience and choosing language that is legible across contexts, by restricting visibility, or by dividing expression across accounts [10]. Teenagers' use of secondary Instagram accounts illustrates a deliberate restoration of context: a smaller audience permits candid expression that would be risky on a polished public profile [11]. The contrast between a main account and a private account therefore should not be read automatically as a split between "false" and "true." It can be a practical strategy for managing multiple legitimate relationships.

Reflected Appraisal and Social Comparison

Cooley's [12] looking-glass self proposes that self-concept develops partly through imagined judgments of others. Digital platforms render this process unusually visible. Likes, comments, views, follower counts, and ranking signals quantify aspects of audience response. These metrics do not provide an objective measure of social worth, but they can be interpreted as evidence of approval or neglect. The psychological effect depends on the meaning assigned to the feedback, the importance of the audience, and the degree to which self-worth is contingent on external evaluation.

Festinger's [13] social comparison theory complements reflected appraisal. People evaluate themselves in relation to available others, especially when objective standards are uncertain.

Social media changes the comparison field by supplying a continuous flow of selectively presented people, including friends, influencers, celebrities, and strangers. Upward comparisons can motivate improvement, but repeated exposure to attractive or successful exemplars may also lower self-evaluation when the comparison target feels relevant and the discrepancy appears difficult to close [14, 15].

Self-Discrepancy and Authenticity

Higgins's [16] self-discrepancy theory distinguishes the actual self from ideal and ought self-guides. A digital profile can function as a representation of an ideal self who a person wishes to be or an ought self who the person believes an audience expects. When the presented identity diverges sharply from lived experience, attention may be repeatedly drawn to the discrepancy. Dejection-related emotions can follow an actual-ideal gap, while anxiety may accompany an actual-ought gap. The model is especially useful because it does not presume that all curation is harmful; distress depends on the type, salience, and perceived attainability of the discrepancy.

Authenticity likewise requires careful definition. It is not equivalent to unfiltered disclosure, nor does it require presenting every aspect of oneself to every audience. In digital settings, authenticity is better understood as experienced congruence between valued aspects of self and chosen expression. Bailey et al. [17] operationalized authenticity as proximity between self-reported personality and personality inferred from Facebook behavior. Across a large observational sample, greater authenticity was associated with higher life satisfaction; a smaller preregistered longitudinal experiment also linked authentic posting with more positive affect within persons. These findings support the value of congruence but do not establish a universal rule for every platform or population.

Technologies of the Self, Networked Identity, and Mediated Selectivity

Foucault's [18] concept of technologies of the self directs attention from presentation to self-formation. Posting, tracking, archiving, and revising are not only messages to others; they are practices through which people observe and work upon themselves. Digital identity is thus constitutive as well as representational. Metrics and memories invite users to monitor the self, while platform norms define which forms of visibility appear desirable. Agency remains present, but it is exercised within an environment of incentives, categories, and unequal visibility.

Turkle [19] emphasized the early internet's capacity for identity experimentation and multiplicity. Her later account warned that continuous connectivity may weaken solitude and promote relationships managed through controlled, editable exchanges [20]. These positions are not contradictory. Digital environments can expand the repertoire of possible selves while also increasing the labor required to maintain them. The key question is whether multiplicity is experienced as exploration and integration or as fragmentation and compulsory availability.

Walther's [21] hyperpersonal model explains why computer-mediated communication can produce highly selective self-presentation and idealized impressions. Reduced cues permit users to allocate attention strategically, while receivers may fill missing information with favorable inferences. Social media adds persistent profiles and public metrics to this process. Nadkarni and Hofmann's [22] review of Facebook motives further suggests that the need to belong and the need for self-presentation operate together. Digital identity is therefore

simultaneously interpersonal and intrapersonal: people seek connection by constructing a version of self that they believe can secure recognition.

Thematic Review

Digital Self-Presentation and Self-Evaluation

Selective presentation is a normal feature of social life, not a pathology created by technology. What changes online is the level of control, persistence, replicability, and quantification. A user can edit a photograph, rehearse a caption, delay a response, and inspect engagement after publication. These affordances make presentation more deliberate, but they also make evaluation more continuous. Zhao et al. [3] found that Facebook identity construction was often implicit and socially anchored: users displayed preferred selves through photographs, affiliations, and relationships rather than through explicit claims. This pattern complicates the idea that digital personas are simply fabricated. They are usually constrained by people who know the user offline, even when selectively enhanced.

The psychological consequences of self-presentation depend on the behavior and the person. A polished profile may help someone articulate goals, consolidate a professional identity, or connect with a valued community. It may also increase self-surveillance if the user begins to view everyday experience through its potential for display. When the imagined audience becomes central, attention shifts from participation in an event to evaluation of how the event will appear. This shift is consistent with Goffman's [9] account of impression management and Hogan's [4] emphasis on persistent exhibitions.

Feedback-seeking is one mechanism through which presentation becomes self-evaluation. Nesi and Prinstein [23] found that social comparison and feedback-seeking on social media were associated with depressive symptoms among adolescents, with effects varying by gender and popularity. Hjetland et al. [24] similarly identified profiles of low, intermediate, and high self-presentation focus. A high focus was associated cross-sectionally with greater symptoms of depression and anxiety, and increases in self-presentation focus were linked longitudinally to small increases in these symptoms. The longitudinal effects were modest, which is important: the study supports psychological relevance without justifying a claim that self-presentation alone produces substantial mental-health deterioration.

Reciprocal effects also remain plausible. Frison and Eggermont [25] found longitudinal associations in both directions between negative Facebook comparison and adolescents' life satisfaction, indicating that unfavorable comparison may reduce satisfaction while lower satisfaction may also increase susceptibility to negative comparison. This bidirectionality is a central reason to avoid interpreting a single cross-sectional association as a one-way media effect.

Reviews reinforce this qualified interpretation. Keles et al. [26] found associations between adolescent social media use and depression, anxiety, and psychological distress, but highlighted methodological weaknesses and mediating factors. Schønning et al. [27] likewise concluded that findings were mixed and that the type and context of use matter. Orben [7] cautioned against broad causal claims based on heterogeneous measures and small associations. Accordingly, the most defensible conclusion is not that identity performance is uniformly harmful, but that approval-oriented and comparison-intensive forms of performance can become risk processes for some users.

Social Comparison, Appearance, and Contingent Worth

Social media supplies comparison targets at a scale and frequency that offline settings rarely match. Chou and Edge [14] found that greater Facebook use, particularly involving less familiar others, was associated with the belief that others had better and happier lives. Vogel et al. [15] linked frequent social media comparison with lower self-esteem and showed experimentally that exposure to upward-comparison profiles could reduce state self-evaluation. These studies illustrate how selective visibility distorts the comparison baseline: the observer compares ordinary private experience with other people's public selections.

Image-centered environments sharpen appearance-based comparison. Chua and Chang's [28] qualitative work with Singaporean teenage girls showed how selfie practices intertwined peer comparison, editing, feedback, and presentation. Mills et al. [29] found that posting selfies, even when participants could retake and edit them, worsened mood and facial satisfaction among young women. The opportunity to edit did not remove the evaluative burden; it may have extended attention to perceived flaws.

Cross-cultural evidence suggests both common mechanisms and culturally variable meanings. Prieler et al. [30] reported relationships among approval-contingent self-worth, appearance comparison on Facebook, and body esteem across European and South Korean samples. The pattern supports the relevance of comparison beyond a single Western setting, but cultural differences caution against treating authenticity, modesty, and public self-enhancement as universal values. The same act omitting a personal opinion, presenting family achievements, or using an edited image may carry different meanings in individualistic and collectivistic contexts.

These findings also clarify why visible metrics matter. A like count can function as social information, reward, and comparison standard at the same time. Yet metrics are shaped by network size, timing, platform distribution, and content visibility; they are not direct measures of relationship quality. Digital literacy should therefore include interpretive literacy: users need not only to know how platforms operate, but also to understand why engagement data are poor proxies for personal worth.

Authenticity, Audience Segmentation, and Identity Integration

Authenticity is frequently positioned as the solution to curated social media, but the concept can itself become performative. A candid photograph, disclosure of vulnerability, or "unfiltered" post may be sincerely expressive, strategically audience-oriented, or both. From a dramaturgical perspective, sincerity and performance are not mutually exclusive. A person can intentionally communicate a genuine value. The more useful distinction is between self-expression that feels congruent and autonomous and expression that feels imposed by anticipated judgment.

Bailey et al. [17] provide the strongest evidence in this corpus that authentic expression can support well-being. Their large observational study and smaller within-person intervention point in the same direction, although the measures and Facebook-centered context limit generalization. Darr and Doss [11] show how audience segmentation can create conditions for authenticity. Teenagers described secondary accounts as more candid because access was limited to trusted peers. This finding extends context-collapse theory: authenticity may require not maximum publicity, but an audience whose norms make particular expressions safe [10].

The authenticity problem is therefore partly architectural. Persistent, public, real-name environments encourage caution, while ephemeral or restricted spaces may enable experimentation. Even an application designed explicitly around spontaneous “realness” does not remove presentation work: BeReal users reported resisting prompts and developing strategies to restore control while simultaneously judging effortless-looking posts as more authentic [31]. Anonymity can permit disclosure of stigmatized identities, but it can also reduce accountability. Neither visibility nor anonymity is inherently authentic. Their effects depend on audience expectations, safety, and the user’s goals.

Ibrahim et al. [32] connect incongruent presentation with impostor feelings. Across several social networking and dating platforms, impostor expression predicted a more adaptable online self, a less authentic self, and a greater sense of freedom online even after accounting for other variables. This combination is revealing: the same person may experience online space as liberating and simultaneously feel phony within it. Digital identity is not governed by a single authenticity continuum. It can increase behavioral freedom while decreasing felt congruence.

Self-discrepancy theory helps interpret this tension [16]. An idealized profile may initially motivate action or allow rehearsal of a hoped-for self. Psychological strain is more likely when the profile becomes an obligation that the person cannot inhabit, or when success depends on maintaining a persona that excludes important experience. A mature digital identity therefore does not require perfect consistency across contexts. It requires enough integration that different presentations remain recognizable as context-sensitive expressions of the same person.

Digital Stress, Availability, and the Archived Self

Digital identity is maintained under conditions of persistent accessibility. Messages can arrive at any time; posts can be evaluated long after publication; and past content can reappear through search, screenshots, and automated memories. Steele et al. [33] conceptualized digital stress as a multidimensional phenomenon that includes availability demands, approval anxiety, fear of missing out, connection overload, and online vigilance. These pressures link identity to time: users are not only concerned with what they present, but also with how quickly they respond and whether they remain socially visible.

The archived self creates a second temporal problem. Adolescence and emerging adulthood ordinarily involve experimentation, revision, and the abandonment of earlier roles. Persistent digital records can detach an old expression from the context in which it was produced. The result may be embarrassment, reputational risk, or a sense that the present self is accountable for every previous version. Hogan’s [4] distinction between performance and exhibition is particularly useful here: the creator may leave the stage, but the exhibit remains available.

At the same time, archives can support narrative continuity. Photographs, messages, and posts allow people to remember relationships and recognize personal change. The psychological effect depends on control and interpretation. An archive chosen and revisited by the user can contribute to autobiographical meaning; an archive surfaced unexpectedly by a platform or hostile audience can feel like a loss of agency.

Broad well-being research supports a mechanism-focused account. Passive consumption and upward comparison are more consistently associated with poorer outcomes than active, reciprocal interaction, although individual responses vary substantially [6, 8]. The need to belong may motivate platform use, while self-presentational motives shape what is shown and

monitored [22]. These pathways can reinforce one another: concern about belonging encourages strategic presentation; uncertain feedback encourages additional monitoring; monitoring increases exposure to comparison; and comparison intensifies concern about one's presentation.

Developmental and Cultural Variation

Adolescents are a central population because identity exploration, peer evaluation, and sensitivity to belonging are developmentally salient. However, describing all young people as vulnerable "digital natives" obscures substantial differences in temperament, social support, platform use, and media literacy. Hjetland et al. [24] found meaningful variation in self-presentation focus, while Nesi and Prinstein [23] observed moderation by gender and popularity. A six-week study of LGBTQ+ youth further showed that identity exploration, context collapse, supportive feedback, and the valence of self-presentation were related in distinct ways; authenticity alone did not predict subsequent self-acceptance or self-esteem [34]. These findings support differential-susceptibility models rather than universal effects.

Adults face different identity tasks. Professional visibility, family roles, political expression, dating, and reputational management may collide within a single searchable identity. Older adults may have a well-established offline self but less familiarity with privacy controls or emerging presentation norms. Researchers should therefore avoid using adolescent findings as a general theory of digital identity.

Cross-cultural limitations remain substantial. Prieler et al. [30] offer evidence that approval contingency and appearance comparison matter across several national contexts, and Chua and Chang [28] show the importance of peer norms in Singapore. Nevertheless, most dominant theories originate in Western accounts of an autonomous individual self. In Georgia, high participation makes digital identity an important local topic [1], but descriptive adoption data cannot substitute for psychological research. Studies should examine how family expectations, language, religion, political polarization, migration, and aspirations toward global culture shape the negotiation of online and offline identities.

Discussion

An Integrated Model of Digital Identity

The reviewed literature supports a three-level model. At the intrapersonal level, digital identity involves self-concept, self-guides, motives, emotion, and self-evaluation. At the relational level, identity is co-produced through audiences, feedback, comparison targets, and group norms. At the platform level, affordances determine persistence, visibility, editability, quantification, and algorithmic distribution. No level is sufficient alone.

Goffman [9] explains the strategic presentation of self, but platform persistence requires Hogan's [4] concept of exhibition. Cooley [12] explains why audience response becomes part of self-concept, while Festinger [13] explains evaluation against available others. Higgins [16] clarifies the affective consequences of gaps among actual, ideal, and ought selves. Foucault [18] shifts attention to repeated practices of self-monitoring and transformation. Walther [21] explains how selective cues and idealized inference develop in mediated interaction, while Turkle [19, 20] keeps both the exploratory promise and the relational cost of networked identity in view.

Together, these theories suggest that digital identity is a feedback system. A person selects content in anticipation of an audience; the platform distributes and quantifies that content; audience responses and comparison information are interpreted; those interpretations alter self-evaluation and future presentation. Bingley et al.'s [5] social identity theory of digital identity strengthens the group-level component of this account by explaining how online and offline memberships organize digital self-definition. The present synthesis complements that contribution by specifying how self-guides, audience feedback, comparison, archival persistence, and algorithmic visibility interact across intrapersonal, relational, and platform levels. The cycle can be constructive when expression is autonomous, relationships are supportive, and feedback is held in perspective. It can become harmful when self-worth is contingent on metrics, comparison is persistently upward, or the represented self becomes an inflexible obligation.

What the Evidence Supports and What It Does Not

Three conclusions are reasonably well supported. First, social media use is psychologically heterogeneous. Reviews do not justify a single direction or large universal effect [6, 7, 26, 27]. Second, specific processes especially upward comparison, feedback-seeking, approval anxiety, and appearance-focused presentation are more informative than aggregate time measures [23, 24, 33]. Third, congruent self-expression and supportive interaction can have positive value [8, 17].

The evidence does not support four common overstatements. It does not show that online and offline identities are always separate; anchored networks often connect them [3]. It does not show that curation is inherently deceptive; all social presentation is selective. It does not establish that social media alone causes a mental-health crisis; bidirectionality and confounding remain plausible. Finally, it does not support treating visible engagement as a valid index of either authenticity or social worth.

The “Black Mirror” as a Conceptual Synthesis

The “black mirror” metaphor integrates these levels without relying on fictional narratives. As a **stage**, the screen enables selective presentation and audience management. As a **social mirror**, it returns reflected appraisals and comparison information. As an **archive**, it preserves earlier performances beyond their original context. As an **algorithmic environment**, it governs which identities become visible and which forms of expression receive measurable reward.

The mirror is “black” not because technology is intrinsically destructive, but because its reflection is partial and difficult to interpret. It combines deliberate expression with behavioral traces, interpersonal judgments with automated classification, and present identity with archived history. The reflected image is therefore neither a transparent self nor a pure fabrication. It is a negotiated representation produced by a person within social and technical constraints.

This metaphor also helps replace the binary of “real” versus “virtual” identity. For many users, digital interaction is part of real social life. The more productive questions are whether an expression is autonomous or compelled, integrated or alienating, supportive or punitive, and open to revision or fixed by persistent records. Digital well-being depends less on eliminating the digital self than on preserving the person’s capacity to interpret, revise, and contextualize it.

Practical Implications

For mental-health practice, assessment should include the client's digital ecology: which platforms are used, for what purpose, before which audiences, and with what emotional aftermath. A general instruction to "use social media less" may miss the mechanism. Interventions may instead address comparison habits, approval-contingent self-worth, compulsive checking, audience boundaries, or the gap between valued identity and public presentation.

For education, digital literacy should extend beyond privacy and misinformation. Students need concepts for understanding selective presentation, imagined audiences, metric ambiguity, context collapse, and the permanence of digital exhibitions. Such knowledge should be paired with emotional and relational skills, because knowing that a feed is curated does not automatically prevent painful comparison.

For platform design, the evidence favors user control over audiences, visibility, metrics, archives, and recommendation settings. Removing all metrics is not the only possible response; meaningful options include private or segmented sharing, friction before impulsive posting, control over automated memories, and explanations of why content is recommended. Design should reduce unnecessary approval pressure without treating users as passive recipients.

Research Priorities

Future research should adopt longitudinal and experience-sampling methods capable of separating selection effects from within-person change. Platform-sensitive measurement is essential: a private messaging exchange, a public short video, and professional networking cannot be represented adequately by a single total-use variable. Studies should measure affordances and activities rather than relying exclusively on brand names, which change over time.

Cross-cultural research should examine whether constructs such as authenticity, idealization, and context collapse have equivalent meanings across cultural settings. More work is also needed with children, midlife adults, older adults, marginalized groups, and users who maintain transnational or multilingual identities. Finally, generative artificial intelligence introduces new questions about authorship and self-representation. When text, images, or avatars are partly generated by automated systems, perceived authenticity may depend on disclosure, control, and whether the output expresses or replaces the user's intentions.

Conclusion

Digital identity is produced through familiar psychological processes operating under novel conditions of persistence, scale, quantification, and algorithmic visibility. In response to the first research question, the central mechanisms are impression management, reflected appraisal, social comparison, self-discrepancy, belonging, selective communication, and repeated self-monitoring. These processes form a feedback loop in which presentation shapes response and response shapes subsequent self-evaluation and presentation.

In response to the second question, social media does not simply increase or decrease authenticity. It changes the conditions under which authenticity is negotiated. Audience collapse and visible metrics can encourage strategic idealization, while audience segmentation

and supportive communities can permit congruent expression. The psychological significance lies in whether users experience control, safety, and alignment with valued aspects of self.

In response to the third question, no single theory is adequate. A combined framework is required: dramaturgy and exhibition explain presentation; the looking-glass self and social comparison explain evaluation; self-discrepancy explains affective tension; technologies of the self explain self-formation and power; and mediated-communication theory explains selective cues and idealized impressions.

In response to the fourth question, the “black mirror” is best understood as a composite metaphor. The screen is a stage, social mirror, archive, and algorithmic environment. Its reflection is consequential but incomplete. A psychologically sustainable digital identity therefore does not require a perfectly unified persona or withdrawal from online life. It requires the capacity to maintain context, revise past expressions, resist reducing self-worth to engagement metrics, and integrate digital experiences into a broader and more flexible sense of self.

References

- [5] Kemp S. *Digital 2026: Georgia*. Report. DataReportal. Singapore, 2025. URL: <https://datareportal.com/reports/digital-2026-georgia>.
- [6] Kemp S. *Digital 2026: Global overview report*. Report. DataReportal. Singapore, 2025. URL: <https://datareportal.com/reports/digital-2026-global-overview-report>.
- [7] Zhao S., Grasmuck S., Martin J. Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, 2008, vol. 24, no. 5, pp. 1816–1836. DOI: <https://doi.org/10.1016/j.chb.2008.02.012>.
- [8] Hogan B. The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society*, 2010, vol. 30, no. 6, pp. 377–386. DOI: <https://doi.org/10.1177/0270467610385893>.
- [9] Bingley W.J., Worthy P., Wiles J., Haslam S.A. A social identity theory of digital identity. *Perspectives on Psychological Science*, 2026, vol. 21, no. 4, pp. 346–382. DOI: <https://doi.org/10.1177/17456916261419813>.
- [10] Kross E., Verduyn P., Sheppes G., Costello C.K., Jonides J., Ybarra O. Social media and well-being: Pitfalls, progress, and next steps. *Trends in Cognitive Sciences*, 2021, vol. 25, no. 1, pp. 55–66. DOI: <https://doi.org/10.1016/j.tics.2020.10.005>.
- [11] Orben A. Teenagers, screens and social media: A narrative review of reviews and key studies. *Social Psychiatry and Psychiatric Epidemiology*, 2020, vol. 55, no. 4, pp. 407–414. DOI: <https://doi.org/10.1007/s00127-019-01825-4>.
- [12] Verduyn P., Ybarra O., Résibois M., Jonides J., Kross E. Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review*, 2017, vol. 11, no. 1, pp. 274–302. DOI: <https://doi.org/10.1111/sipr.12033>.
- [13] Goffman E. *The presentation of self in everyday life*. Garden City, New York: Doubleday, 1959.
- [14] Marwick A.E., boyd d. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 2011, vol. 13, no. 1, pp. 114–133. DOI: <https://doi.org/10.1177/1461444810365313>.

-
- [15] Darr C.R., Doss E.F. The fake one is the real one: Finstas, authenticity, and context collapse in teen friend groups. *Journal of Computer-Mediated Communication*, 2022, vol. 27, no. 4, article zmac009. DOI: <https://doi.org/10.1093/jcmc/zmac009>.
- [16] Cooley C.H. *Human nature and the social order*. New York: Charles Scribner's Sons, 1902.
- [17] Festinger L. A theory of social comparison processes. *Human Relations*, 1954, vol. 7, no. 2, pp. 117–140. DOI: <https://doi.org/10.1177/001872675400700202>.
- [18] Chou H.-T.G., Edge N. "They are happier and having better lives than I am": The impact of using Facebook on perceptions of others' lives. *Cyberpsychology, Behavior, and Social Networking*, 2012, vol. 15, no. 2, pp. 117–121. DOI: <https://doi.org/10.1089/cyber.2011.0324>.
- [19] Vogel E.A., Rose J.P., Roberts L.R., Eckles K. Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture*, 2014, vol. 3, no. 4, pp. 206–222. DOI: <https://doi.org/10.1037/ppm0000047>.
- [20] Higgins E.T. Self-discrepancy: A theory relating self and affect. *Psychological Review*, 1987, vol. 94, no. 3, pp. 319–340. DOI: <https://doi.org/10.1037/0033-295X.94.3.319>.
- [21] Bailey E.R., Matz S.C., Youyou W., Iyengar S.S. Authentic self-expression on social media is associated with greater subjective well-being. *Nature Communications*, 2020, vol. 11, no. 1, article 4889. DOI: <https://doi.org/10.1038/s41467-020-18539-w>.
- [22] Foucault M. Technologies of the self. In: *Technologies of the self: A seminar with Michel Foucault*. Edited by L.H. Martin, H. Gutman, and P.H. Hutton. Amherst, Massachusetts: University of Massachusetts Press, 1988, pp. 16–49.
- [23] Turkle S. *Life on the screen: Identity in the age of the Internet*. New York: Simon & Schuster, 1995.
- [24] Turkle S. *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books, 2011.
- [25] Walther J.B. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 1996, vol. 23, no. 1, pp. 3–43. DOI: <https://doi.org/10.1177/009365096023001001>.
- [26] Nadkarni A., Hofmann S.G. Why do people use Facebook? *Personality and Individual Differences*, 2012, vol. 52, no. 3, pp. 243–249. DOI: <https://doi.org/10.1016/j.paid.2011.11.007>.
- [27] Nesi J., Prinstein M.J. Using social media for social comparison and feedback-seeking: Gender and popularity moderate associations with depressive symptoms. *Journal of Abnormal Child Psychology*, 2015, vol. 43, no. 8, pp. 1427–1438. DOI: <https://doi.org/10.1007/s10802-015-0020-0>.
- [28] Hjetland G.J., Finserås T.R., Sivertsen B., Colman I., Hella R.T., Andersen A.I.O., Skogen J.C. Digital self-presentation and adolescent mental health: Cross-sectional and longitudinal insights from the "LifeOnSoMe"-study. *BMC Public Health*, 2024, vol. 24, no. 1, article 2635. DOI: <https://doi.org/10.1186/s12889-024-20052-4>.
- [29] Frison E., Eggermont S. "Harder, better, faster, stronger": Negative comparison on Facebook and adolescents' life satisfaction are reciprocally related. *Cyberpsychology, Behavior, and Social Networking*, 2016, vol. 19, no. 3, pp. 158–164. DOI: <https://doi.org/10.1089/cyber.2015.0296>.
-

- [30] Keles B., McCrae N., Grealish A. A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, 2020, vol. 25, no. 1, pp. 79–93. DOI: <https://doi.org/10.1080/02673843.2019.1590851>.
- [31] Schønning V., Hjetland G.J., Aarø L.E., Skogen J.C. Social media use and mental health and well-being among adolescents: A scoping review. *Frontiers in Psychology*, 2020, vol. 11, article 1949. DOI: <https://doi.org/10.3389/fpsyg.2020.01949>.
- [32] Chua T.H.H., Chang L. Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. *Computers in Human Behavior*, 2016, vol. 55, pp. 190–197. DOI: <https://doi.org/10.1016/j.chb.2015.09.011>.
- [33] Mills J.S., Musto S., Williams L., Tiggemann M. "Selfie" harm: Effects on mood and body image in young women. *Body Image*, 2018, vol. 27, pp. 86–92. DOI: <https://doi.org/10.1016/j.bodyim.2018.08.007>.
- [34] Prieler M., Choi J., Lee H.E. The relationships among self-worth contingency on others' approval, appearance comparisons on Facebook, and adolescent girls' body esteem: A cross-cultural study. *International Journal of Environmental Research and Public Health*, 2021, vol. 18, no. 3, article 901. DOI: <https://doi.org/10.3390/ijerph18030901>.
- [35] Pinch A., Fiers F., Birnholtz J., Fisher J., Reilly B. "Is it time for me to be authentic?": Understanding, performing, and evaluating authenticity on BeReal. *New Media & Society*, 2025, vol. 27, no. 11, pp. 6315–6333. DOI: <https://doi.org/10.1177/14614448241267731>.
- [36] Ibrahim F., Herzberg P.Y., Stöven L.M. Feeling phony online: The impostor phenomenon's link to online self-presentation, self-esteem, and social network site use. *Acta Psychologica*, 2024, vol. 248, article 104342. DOI: <https://doi.org/10.1016/j.actpsy.2024.104342>.
- [37] Steele R.G., Hall J.A., Christofferson J.L. Conceptualizing digital stress in adolescents and young adults: Toward the development of an empirically based model. *Clinical Child and Family Psychology Review*, 2020, vol. 23, no. 1, pp. 15–26. DOI: <https://doi.org/10.1007/s10567-019-00300-5>.
- [38] Chen Y.A., Mares M.-L. Social media self-presentation of LGBTQ+ youth in the United States: The role of identity exploration, context collapse, and supportive feedback. *Journal of Computer-Mediated Communication*, 2025, vol. 30, no. 6, article zmaf020. DOI: <https://doi.org/10.1093/jcmc/zmaf020>.

Article received 2026-08-01