

The Reader's Digital Shadow: A Hermeneutic Inquiry into the Formation of the Self Through Reading Traces

Dongchun Chen

School of Foreign Studies, Guangdong University of Finance and Economics

Author Note

I have no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Dongchun Chen, School of Foreign Studies, Guangdong University of Finance and Economics, Guangzhou, Guangdong, China.

Email: chendongchun@gdufe.edu.cn

Abstract

This article investigates the formation of the digital self through a hermeneutic analysis of digital reading traces. We posit the concept of the “digital shadow” to theorise the aggregate of passive, behavioral data (e.g., reading duration, highlighting patterns) generated unintentionally during platformized reading. This shadow is not merely a residue but an active, hermeneutic entity. We argue that self-formation occurs through a cyclical process: first, in the fusion of horizons (Gadamer, 2004, p.301-302) between reader and text, which produces discrete traces; second, as platforms algorithmically curate these traces into a coherent narrative profile—a mirror of the reader’s intellectual habits; and finally, as the reader interprets this algorithmic reflection of their own digital shadow, thereby integrating it into their ongoing self-understanding. This study provides a framework for understanding how reading, a fundamental act of interpretation, becomes a datafied site for the mediated construction of the intellectual self.

Keywords: digital Shadow, reading traces, hermeneutics, platformized reading, algorithmic curation, intellectual self, Gadamer

The proliferation of digital reading platforms—from Kindle and Apple Books to WeChat Read and Goodreads—has fundamentally transformed the ancient practice of reading. This transformation extends beyond mere convenience, portability, or access; it signifies a profound shift in the very phenomenology of reading, turning a traditionally private, contemplative act into a practice of pervasive data generation. Every interaction with a digital

text—a highlight, a note, a pause on a page, or even the decision to abandon a book—is captured, quantified, and stored as a digital trace. While scholarship in digital sociology and media studies has effectively mapped the political economy of these platforms and their algorithmic governance (van Dijck, 2014, p. 199; Striphas, 2009, p. 135), a critical ontological dimension remains underexplored: how does the datafication of reading—a core hermeneutic process of meaning-making—reconfigure the formation of the intellectual self? Existing literature on the “quantified self” (Lupton, 2016, p. 46) often focuses on health and fitness, while studies on digital footprints primarily address issues of privacy and surveillance. However, the specific ways in which passive behavioral data from reading are curated into a narrative of the self demand a more nuanced theoretical framework. This gap is particularly evident when considering how platforms algorithmically synthesize these discrete traces into coherent profiles (e.g., “Your Year in Review,” “Popular Highlights”) and reflect them back to the user, thus creating a new object for self-interpretation.

This paper addresses this gap by introducing the concept of the “digital shadow”—defined as the cumulative, datafied representation of a user’s unintentional and passive reading behaviors, systematically collected as a byproduct of platform engagement. Moving beyond the concept of active self-presentation, the digital shadow represents a form of passive identity construction where the self is authored not only by the user but also by the algorithmic processing of their behavioral remnants. To theorize this process, we turn to the philosophical hermeneutics of Hans-Georg Gadamer, specifically his concept of the “fusion of horizons” (*Horizonverschmelzung*:). Gadamer’s framework, which posits that understanding is always a dialogical event between the interpreter’s pre-understandings and the meaning of the text, provides a powerful lens for analyzing the tripartite dialogue between the reader, the text, and the platform’s algorithms. In this mediated hermeneutic circle, the platform functions as an

algorithmic other that actively curates and interprets the reader's digital shadow, presenting it back as a mirror for self-contemplation.

Consequently, the central research question guiding this inquiry is: How does the algorithmic mediation and curation of digital reading traces (the “digital shadow”) facilitate a new hermeneutic process of self-formation, and what are the implications for our understanding of agency and identity in the platformized lifeworld? This study argues that the intellectual self is increasingly formed through a cyclical process of interpreting this algorithmically pre-processed mirror, blurring the line between the narrating “I” and the narrated “me.” Through a hermeneutic analysis of features like algorithmic recommendations and personalized reading reports, this paper aims to develop a critical framework for understanding reading no longer solely as an encounter with texts, but as a datafied dialogue with one's own digitally reflected self.

1. Theoretical Framework: Narrating the Self in the Hermeneutic Circle

The conceptualization of the digital self within this inquiry necessitates a theoretical departure from models that posit the self as a stable entity awaiting representation or as a mere effect of structural determination. Instead, this study posits the digital self as a dynamic, relational process—an ongoing event of interpretation and becoming that is profoundly mediated by the technical milieu. This processual view finds its most potent articulation in the philosophical hermeneutics of Hans-Georg Gadamer, whose framework provides the conceptual architecture for understanding how the digital shadow of reading traces becomes integral to self-formation. Gadamer's central thesis—that understanding is never a solitary mirroring of a pre-given world but a dialogical fusion of horizons (*Horizonverschmelzung*) (Gadamer, 2004, pp. 301-302)—offers a critical lens for dissecting the tripartite dialogue constitutive of the platformized reader. For Gadamer, a horizon is the spectrum of presuppositions, prejudices (*Vorurteile*) (Gadamer, 2004, p. 305), and cultural contexts that

precondition any act of understanding. The fusion occurs when the horizon of the interpreter engages with the horizon of the text (Gadamer, 2004, p. 307), not to overcome their difference, but to generate a new, expanded understanding.

This study transposes this model onto the digital reading environment, arguing that the hermeneutic circle is now technologically extended and involves multiple agents. The initial fusion occurs between the reader (with their unique horizon of pre-understandings) and the digital text, an engagement that materially produces the digital trace—a highlight or note being the crystallization of this interpretive moment. However, this trace is immediately appropriated by the platform’s algorithms, which operate according to their own horizon of calculability—a logic predicated on pattern recognition, correlation, and prediction. The platform’s agency lies in its capacity to curate these dispersed traces, assembling them into a new composite text: the data-driven narrative profile or the digital shadow. This algorithmic curation is not neutral; it is an act of interpretation that imposes a specific, often reductive, logic of coherence and value upon the reader’s intellectual life.

The pivotal moment of self-formation, therefore, occurs in the subsequent fusion—the reader’s encounter with this algorithmically generated mirror. Here, the reader is no longer simply interpreting the primary text but is engaged in a meta-hermeneutic act: interpreting the platform’s interpretation of their own reading self. The reader’s horizon fuses with the algorithmic horizon embedded in the curated digital shadow (e.g., a personalized recommendation list or an annual reading report). This fusion can validate, alienate, or reshape the reader’s self-conception, effectively closing a feedback loop where the self is continually refined through dialogue with its own datafied double. Thus, Gadamer’s dialogical model allows us to reframe the research question: it is no longer simply about how data is collected, but about how understanding itself is structured within this techno-hermeneutic circuit. It provides the vocabulary to analyze how agency is distributed between human and non-human

interpreters and how the intellectual self emerges from this ongoing, mediated conversation between the reader, the text, and the algorithmic other.

2. The Architecture of the Shadow: The Reading Platform as Narrative Engine

The formation of the reader's digital shadow is not an ethereal process but is materially grounded in the specific architectural affordances and computational operations of the digital reading platform. This platform functions as a hermeneutic engine, a system designed to translate the fluid, contextual act of reading into stable, interoperable data, and then to re-translate that data back into a narrative that can act upon the reader. This translation occurs through a layered process that can be conceptualized through the fused lens of Gadamer and Ricoeur (1991), moving from the pre-reflective engagement to the configured narrative. The architecture begins with *Mimesis*₁ (Prefiguration), where the platform's very design—its highlighting tool, its social features, its seamless connectivity—establishes the realm of possible interactions. It prefigures what counts as a valuable trace: a tap to highlight is privileged over a pondering gaze out a window; a shareable note is valued over a private, ephemeral thought. This technical *a priori* creates a horizon of expectations that shapes reading practice before a single word is consumed, channeling the reader's hermeneutic activity into data-generating pathways (cf. Ricoeur, 1991, pp. 21-28).

The core of the engine lies in *Mimesis*₂ (Configuration), where the platform's algorithmic logic performs the act of employment. Disparate traces—each a Gadamerian moment of fusion between a reader and a text—are extracted from their native context. The highlight is severed from the paragraph that prompted it; the reading speed is dislocated from the complexity of the prose. Through operations of clustering, pattern recognition, and correlation, these decontextualized data points are woven into new, algorithmic plots. A narrative profile emerges: "The reader who lingered on this page also bought..."; "Your reading habits align with scholars of..."; "Your October was dedicated to existentialist philosophy."

This is the digital shadow crystallized—a story configured not by the reader’s reflective consciousness, but by the platform’s computational interpretation of their behavior. It is a narrative authored by the algorithmic other, imbued with its values of quantifiability, comparability, and predictive utility.

Finally, this configured narrative is re-injected into the reader’s lifeworld through the interface in Mimesis₃ (Refiguration). The platform becomes a stage for a profound, meta-hermeneutic encounter. The reader is presented with their own data-driven double: the Spotify Wrapped-style “Year in Review,” the personalized recommendation list that claims to know “what you’re like.” This moment constitutes a second-order fusion of horizons. The reader’s lived sense of self (their Gadamerian horizon) now engages with the narrative horizon of their digital shadow. This encounter is not passive; it demands interpretation. Does the reader recognize themselves in this portrait? Do they accept this algorithmic biography as authoritative, or do they resist it? This moment of refiguration completes the hermeneutic circuit, but it also potentially alters the reader’s future pre-understandings (Mimesis₁ again). A reader who internalizes the platform’s label of them as a “fast reader” may begin to value speed over depth; one who is consistently recommended popular bestsellers may have their literary horizons narrowed rather than expanded. Thus, the architecture of the shadow does not merely report on the self; it actively participates in its re-creation, turning the practice of reading into an ongoing dialogue between the human seeker of meaning and the machine producer of narratives.

3. Case Analysis: The Hermeneutic Circle in Action

This section empirically examines the theoretical model of the “dual hermeneutic circle” through a critical analysis of specific features endemic to contemporary digital reading platforms. The cases are selected based on their conceptual prevalence and their demonstrable role in the algorithmic mediation of the reading experience. The analysis focuses on the

Amazon Kindle ecosystem, which serves as a paradigmatic case due to its market dominance, sophisticated data-capture capabilities, and explicit curation features. This methodological focus allows for a deep, contextualized inquiry into the mechanisms of platform-based hermeneutics.

The platform's operation can be theorized as a hermeneutic process in itself, wherein raw behavioral data is transformed into a curated narrative of the self. This process is underpinned by a specific architecture of datafication, which translates qualitative reading experiences into quantifiable data points that form the constitutive elements of the digital shadow. The analysis for each case will draw upon the typical data points collected, which we synthesize in the table below. This "grammar of the shadow" forms the raw material for algorithmic curation. The data points are not neutral metrics but carry inherent hermeneutic significance.

Table 1: The Grammar of the Digital Shadow

Data Point	Description	Hermeneutic Significance
Progress (%)	The percentage of the text completed.	A metric of engagement and commitment; infers interest level and reading pace.
Time Spent	Duration spent on a page, chapter, or entire book.	Distinguishes deep reading from skimming; identifies difficult or engaging sections.
Highlights & Notes	Text selections marked by the user and any accompanying notes.	The most explicit signal of interpretive engagement. Reveals what the reader finds meaningful, surprising, or noteworthy. The content of notes provides direct semantic data.
Popular Highlights	Passages highlighted frequently by other users.	Creates a social horizon of pre-understanding; signals collectively validated meaning.
Purchase/Download History	The library of books a user has acquired.	Defines the user's explicit intellectual territory and potential fields of interest.

Data Point	Description	Hermeneutic Significance
Search & Browse History	Queries within the platform and books viewed but not purchased.	Reveals intentionality, curiosity, and exploratory paths not yet realized.

3.1 Case 1: The Socialization of Interpretation through “Popular Highlights”

The “Popular Highlights” feature is a quintessential example of how platforms inject a social horizon into the hermeneutic process. This feature operationalizes the highlights data point (Table 1) by applying a quantitative threshold to create a new, socially validated textual paratext. This constitutes a form of algorithmically mediated pre-understanding, where the collective judgment of past readers is presented as a pre-given framework for interpreting the text.

Consider a reader encountering the passage, “Perhaps one did not want to be loved so much as to be understood,” in George Orwell’s 1984. Without mediation, the passage’s meaning is negotiated privately between reader and text. However, when annotated as a “Popular Highlight” (e.g., “Highlighted by 5,243 readers”), the interpretive act is immediately socialized. The reader’s horizon fuses not only with the text but with the quantified horizon of the crowd. This can lead to the standardization of interpretation, where the algorithmic emphasis on consensus marginalizes idiosyncratic or critical readings (van Dijck, 2013, p. 154). The reader’s agency lies in their critical response to this socially quantified pre-understanding—whether they accept, reject, or simply notice its influence.

3.2 Case 2: The Construction of Projected Identity through Personalized Recommendation

Personalized recommendation systems (e.g., “Because you read...”) function as algorithmic bibliographers. They synthesize a wide array of data points—including purchase

history, reading progress, time spent, and highlights—to infer user affinity and construct a projective identity for the reader. This is not a reflection of a pre-existing self but an extrapolation of a datafied identity into a future trajectory. The Self-Reflection creates a pathway for further self-formation, effectively shaping the reader's horizon of future intellectual possibilities.

A user's deep engagement with Yuval Noah Harari's *Sapiens* (evidenced by slow reading pace and extensive notes on chapters concerning the Cognitive Revolution) signals to the algorithm an interest in macro-historical narratives. The platform may subsequently recommend Jared Diamond's *Guns, Germs, and Steel*. This recommendation is productive, opening new avenues for exploration. However, it also performs a hermeneutic foreclosure. The same user might have a nascent interest in poetry, but because this interest lacks a robust data signature, the algorithm is unlikely to recommend the works of Mary Oliver. Thus, the algorithmic mirror reflects and reinforces a specific, data-rich facet of the self, while leaving other potential intellectual identities underdeveloped and unreflected.

3.3 Case 3: The Externalization of Memory in “Year in Review” Reports

Platform-generated summaries (e.g., “Your Year in Reading”) represent the most advanced form of algorithmic curation, where the platform assumes the role of a biographer of the self. These reports perform a reification of the intellectual self, distilling a year of qualitative experiences into a shareable narrative built from quantifiable metrics. This forces the user into a position of second-order self-hermeneutics, where they are tasked with interpreting the platform's interpretation of their own past.

A report might celebrate a user for reading 30 books and highlight *Atomic Habits* as their “most highlighted” book. This narrative promotes an identity of productivity and self-optimization. However, this curated story may obscure the phenomenological truth of the reading experience. The user might have found deep value in a complex novel like Richard

Powers's *The Overstory*, which they read slowly and reflectively, resulting in fewer highlights. The report's algorithmic logic, privileging quantity and explicit engagement, cannot capture this qualitative depth. This can create hermeneutic dissonance, a disconnect between the data-driven narrative and the user's lived experience, raising fundamental questions about the authority to narrate one's intellectual life.

3.4 Concluding Synthesis: The Imperative for Critical Hermeneutics

The cases collectively demonstrate that the digital shadow is not a passive byproduct but an active, constitutive agent in the formation of the contemporary self. The platform's algorithmic curation creates a powerful feedback loop within the hermeneutic circle, shaping the very pre-understandings that inform future reading and self-interpretation. The critical question that emerges is not whether this mediation occurs, but how we engage with it critically. This analysis underscores the imperative for developing a critical hermeneutics of datafication—a reflective practice where users learn to interrogate the algorithmic mirrors presented to them, understanding their constructed nature and the interests they serve. This conclusion provides a direct conceptual bridge to the discussion of agency, ethics, and resistance in the following section.

4. Discussion: Agency, Ethics, and the Mediated Intellectual Self

The empirical analysis in the previous section confirms the operation of the dual hermeneutic circle and reveals the digital shadow as a constitutive agent in self-formation. This phenomenon echoes earlier insights into how technology serves as a mirror for the self. Sherry Turkle's (1984, p51) seminal work, *The Second Self*, pioneered this line of inquiry, arguing that computers act as a "second self"—a projective surface that shapes and reflects our identities. Turkle's focus was on the psychological interplay between individuals and stand-alone machines. This study extends this foundational concept into the platformized, algorithmic age, arguing that the "digital shadow" is the new, datafied incarnation of the "second self," now

dynamically constructed and curated by platforms rather than subjectively projected onto a static machine.

4.1 The Paradox of Agency: Curated and Curating

The central paradox revealed by this inquiry lies in the reconfiguration of readerly agency. The platform does not simply erase agency; it re-distributes and re-orient it.

On one hand, the platform's architecture creates an illusion of passivity by channeling user agency into reactive choices within algorithmically curated pathways. The reader is positioned as both curator and curated. In the primary hermeneutic circle (reader-text), agency is exercised through active interpretation, highlighting, and note-taking. However, in the metacircle (reader-mirror), the reader's agency is often channeled into reactive choices: to accept or reject a recommendation, to identify with or feel alienated by a yearly report. This creates an illusion of passivity, where the user feels in control of their choices while navigating a pathway pre-structured by algorithmic logic. Agency is reduced to selection within a curated set of options, a form of "managed autonomy" that serves the platform's economic ends.

On the other hand, algorithmic mediation can be productively enabling by expanding hermeneutic horizons, yet it ultimately constitutes a form of managed autonomy that serves platform ends. The algorithmic bibliographer can open doors to intellectual domains the reader might not have discovered independently, effectively expanding their hermeneutic horizon. The key question, therefore, is not whether agency exists, but what form it takes and who ultimately benefits from its exercise. The model of the dual hermeneutic circle allows us to map this complex interplay between the user's interpretive actions and the platform's curatorial power, revealing a relationship that is neither purely oppressive nor purely liberating but is fundamentally ambivalent.

4.2 Ethical Implications: Interpretive Authority, Intellectual Privacy, and Bias

The technological mediation of self-formation through reading raises urgent ethical

questions that strike at the core of intellectual autonomy and human dignity.

First, the shift of interpretive authority from the reader-text dyad to opaque algorithmic processes represents a fundamental redistribution of epistemic power. The analysis challenges the traditional locus of interpretive authority. In Gadamer's model, authority is earned through recognized knowledge and is always open to critical questioning. In the platform ecosystem, authority is assigned to algorithmic processes whose logic is often proprietary, opaque, and optimized for engagement, not truth or depth. When a platform tells a reader what is "important" in a text (via Popular Highlights) or who they "are" as a reader (via annual reports), it exercises a form of authority that is difficult to contest because its mechanisms are hidden. This constitutes a significant shift in epistemic power, from the reader-text dyad to the platform-as-mediator.

Secondly, the datafication of reading traces constitutes a profound violation of intellectual privacy, commodifying the intimate sphere of cognitive engagement. Reading traces constitute a form of intellectual privacy—a record of our curiosity, doubts, struggles, and moments of transcendence. The datafication and curation of this interiority represent a profound intrusion. Unlike a social media post, which is performative by nature, reading data is often an unintentional byproduct of deep cognitive engagement. Its capture and use for profiling and profit without explicit, informed consent violates a fundamental boundary of the self. The digital shadow is, in essence, a datafied unconscious, and its commercial exploitation raises serious concerns about the commodification of human consciousness.

Thirdly, algorithmic bias is critical. The hermeneutic process within the platform is not neutral. Algorithms are trained on data that reflects existing cultural and economic biases. This can lead to a hermeneutic injustice, where the algorithmic mirror systematically misrecognizes or devalues certain types of reading or intellectual pursuits. For instance, a system optimized for bestsellers may fail to adequately support or recognize the value of slow, scholarly, or non-

Western reading patterns, thereby reinforcing existing cultural hegemonies under a guise of technological neutrality.

4.3 Towards a Critical Hermeneutics: Resistance and Self-Reflection

In light of these challenges, a passive stance is untenable. The conclusion must be a call for the development of a critical hermeneutics of the self in the digital age.

Hermeneutic resistance requires conscious critical detachment from the algorithmic mirror and tactical engagement to disrupt its curated narratives. This involves becoming conscious of the metacircle itself. It requires users to adopt a stance of critical detachment towards the algorithmic mirror, constantly asking: “Why is this being shown to me? What story is being told about me, and by whom? What is being emphasized and what is omitted?” Tactics could include deliberately diversifying reading patterns to “confuse” the algorithm, refusing to engage with certain curated features, or using alternative, less intrusive platforms. This is not a rejection of technology but a demand for a more self-aware and negotiated relationship with it.

Moreover, the ultimate goal is to harness this critical awareness for richer self-understanding. The algorithmic mirror, for all its flaws, can serve as a provocation for self-inquiry. The dissonance felt when disagreeing with a “Year in Review” report is not an endpoint but a starting point for a deeper dialogue with oneself: “If that data-story isn’t right, then what is my story?” In this way, the mediated self-formation process can be productively subverted. By critically interrogating the platform’s interpretation, we may actually arrive at a more nuanced, authentic, and defensible self-narrative—one that is achieved not through passive reception but through active, critical resistance and reflection. This reclaims agency not by escaping the metacircle, but by turning it into a catalyst for a more profound hermeneutics of the self.

5. Conclusion

This study has argued that the phenomenon of platformized reading necessitates a fundamental expansion of our hermeneutic vocabulary. We have moved beyond the classic dyad of reader and text to confront a triadic structure involving the reader's digital shadow—the aggregated, datafied trace of their engagement. Through a synthesis of hermeneutic theories, we have developed a dual hermeneutic circle model to articulate self-formation in this context. The first circle, between reader and text, remains foundational to understanding, while the second circle emerges through the platform's algorithmic curation of traces into a coherent narrative mirror. Our central argument identifies the interpretation of this algorithmic mirror as a distinct and powerful moment in the hermeneutics of the self, establishing the digital shadow not as an epiphenomenon but as a constitutive agent that actively shapes the modern intellectual self by providing a new object for interpretation. The primary contribution of this article lies in developing a robust hermeneutic framework for critically analyzing the datafication of intellectual life. By introducing the concepts of “digital shadow” and “algorithmic mirror,” we provide a nuanced vocabulary that captures both the passive and active dimensions of this phenomenon. Our model bridges classic hermeneutic theory with critical algorithm studies, demonstrating how the hermeneutic circle is technologically mediated and reconfigured in the digital age. This framework enables a shift in discussions about digital reading from concerns over distraction or superficiality to a more fundamental inquiry into the politics and poetics of self-formation—examining who gets to interpret us and how those interpretations guide our future becoming. This theoretical inquiry opens several promising avenues for future research, including empirical validation through qualitative studies with platform users, comparative analysis of different reading platforms' curatorial grammars, application of the framework to other domains of datafied selfhood such as fitness tracking or music streaming, and normative design-oriented research into what hermeneutically just platforms might look like and how to create interfaces that foster critical self-reflection rather than passive consumption.

References

- Gadamer, H.-G. (2004). *Truth and Method* (2nd rev. ed., J. Weinsheimer & D. G. Marshall, Trans.). Continuum.
- Lupton, D. (2016). *The Quantified Self: A Sociology of Self-Tracking*. Polity Press.
- Ricoeur, P. (1991). *From Text to Action: Essays in Hermeneutics, II* (K. Blamey & J. B. Thompson, Trans.). Northwestern University Press.
- Striphas, T. (2009). *The Late Age of Print: Everyday Book Culture from Consumerism to Control*. Columbia University Press.
- Turkle, S. (1984). *The Second Self: Computers and the Human Spirit*. Simon & Schuster.
- Van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press.