Proceedings of the 5th International Conference on Art, Design and Social Sciences
DOI: 10.54254/2753-7064/2026.39390

Visual Hegemony and the Crisis of Literature in the Era of "AI Text-to-Image"

Qingyu Yang

School of Humanities, Jiangnan University, Wuxi, China
yyyqqqyyy@foxmail.com

Abstract. The technology of "AI text-to-image" has driven a leap in visual production methods and formed a new type of "visual hegemony" within a specific media ecology, posing multiple crises to traditional literature, which is grounded in deep metaphor, subjective spirit, and temporal experience. The default logic of AI-generated images often compresses the open meaning-space of literary metaphor; algorithmic recommendation and instant visual feedback reshape people's attentional structure and affect the narrative, comprehension, and even aesthetic capacities of the subjects of literary activity; the convenient mode of text-to-image sets the instant, concrete, and massive AI images in competition with the diachronic cultural memory carried by literature.The fundamental contradiction lies between two regimes: one that privileges efficiency, clarity, and calculability, and the other that cherishes ambiguity, polysemy, and ineffability. Yet danger and opportunity go hand in hand; the same tools can facilitate cross-media expression and prompt literature to reaffirm its own irreplaceability.

Keywords: AI text-to-image, visual hegemony, literary crisis

1. Introduction

In his seminal work *The Society of the Spectacle*, Guy Debord declares: "In the society where the conditions of modern production are ubiquitous, life itself appears as a vast accumulation of spectacles." [1] Yet in the age of artificial intelligence, the spectacle is no longer merely an accumulation, but is increasingly liable to be generated by AI on demand and in real time. Built on large-scale text-image pre-training models and generating images through natural-language prompts, the technology of "AI text-to-image" is profoundly reshaping our visual environment, cognitive framework, and cultural logic. It accelerates the production of "spectacles" and pushes us into the order of "simulacra" warned of by Jean Baudrillard. Thus, the standardized, decontextualized, and quickly consumable visuality supported by this specific technical logic is increasingly becoming the dominant paradigm for organizing experience, representing reality, and even defining the "real." Literature's important position and unique aesthetic value within the cultural ecology face the risk of being marginalized, flattened, and instrumentalized.

As defined in this paper, visual hegemony refers to the priority obtained by the visual at the cognitive, cultural, and even institutional levels. It should be noted that the "AI text-to-image" discussed in this paper refers specifically to the typical mode of use within an ecosystem dominated

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7064/2026.39390

by platformization, commercialization, and algorithmic recommendation. Against this background, the power relation between literature and image is further upgraded: it is no longer a matter of how language describes images, or how images interpret texts; rather, it is a matter of how algorithms, on the basis of language as prompt words, directly requisition and define the visual incarnation of texts. This paper attempts to analyze, from the three levels of the mechanism of meaning, the mechanism of the subject, and the mechanism of cultural memory, the influence of AI text-to-image on literature and its limits.

2. Algorithmic synecdoche and the collapse of the metaphorical space

In the default ecology dominated by platformization and algorithmic recommendation, the operation of "AI text-to-image" technology tends to replace the mechanism of "metaphor," on which literature depends for its survival, with an efficient yet impoverished "algorithmic synecdoche" — that is, bypassing the reader's deep interpretation and mental-image construction, it rapidly converts the literary text into a set of visual symbols.

The "structure of appeal" spoken of by the reception-aesthetician Iser [2] means that literary meaning depends on the reader's participation in co-construction through life experience and association. The operative mechanism of "metaphor" [3] lies precisely in connecting two seemingly distant and heterogeneous objects, thereby stimulating the reader's new meaning-association and emotional resonance amid the subtle differences of semantics, and opening up an infinite interpretive space that requires the reader to fill it with imagination and personal experience. For instance, the verbal description "Before my bed the bright moonlight shines, I take it for frost upon the ground" evokes the reader's rich imagination, and the train of thought involuntarily connects with the poet's unique individual experience, emotional projection, and cultural imagination. The kind of "metaphor" abundantly adopted in literature guards the ambiguity, depth, and spirituality of meaning.

Yet what "AI text-to-image" technology is best at handling is superficial, pattern-based association. Borrowing the concept of "synecdoche" from rhetoric, this paper proposes the analytic term "algorithmic synecdoche" to describe a default tendency: using statistically salient, local, and fragmented visual symbols to substitute for a more complex, integral meaning-whole. This is not to say that AI generation can only ever be synecdoche in essence, but that under the combined force of platform optimization and popular expectation, AI images tend to use the most salient and common visual "parts" to directly replace the complex meaning-space, directly "filling" the literary "blank" that awaits "filling." By learning statistical correlations in massive text-image data, it has built a global "visual synecdoche dictionary." For example, when a user inputs text from Kafka's *The Metamorphosis* for image generation, what AI outputs is often not a philosophical visualization of alienation and absurdity, but more likely a collage of database labels such as "beetle" and "room." It should be pointed out that, although professionals such as book illustrators have long visualized literary texts, AI far surpasses traditional illustration in speed, scale, and reproducibility, and its default logic relies more on statistical co-occurrence; and, further, the artistic practice of a painter painting also carries the interpretive-adaptive flavor of the reader's "re-creation." Therefore, the difference between AI text-to-image and traditional illustration is not merely a difference in efficiency, but a difference in the mode of meaning production.

Metaphor is productive; it requires the reader's mental participation to accomplish the leap of meaning. Synecdoche is substitutional; it provides ready-made answers and closes off the space of imagination. The real-time visualization of "AI text-to-image" compresses the originally open and profound world of metaphor into a vast but closed visual database, leveling the "blank" and "unfilled

DOI: 10.54254/2753-7064/2026.39390

point" inherent in the literary text, and replacing them with a certain visual finalization. Of course, readers may still re-interpret, resist, or read image and text intertextually; but in the default path, they often merely passively degrade from active co-constructors of meaning into acceptors of visual results.

This means that literary reception is compressed from a "temporal" process requiring deep investment and repeated savoring into a "spatial" consumption that is instantly presented and readily grasped at a glance. What is lost from this may be not merely a certain literary experience, but the complex, ambiguous, and profound human spiritual world constructed by language.

3. Data discipline and the reshaping of the literary subject

Michel Foucault's theory of "discipline" [4] reveals the power mechanism by which modern society shapes the individual into a "docile body." Borrowing this analytic perspective, once "AI text-to-image" forms a joint force with algorithmic recommendation, it is no longer merely an isolated image-generation tool, but becomes a digital disciplinary mechanism that shapes reading patience, creative expectation, and aesthetic habit. This may cause the reading subject to lose the patience and ability to appreciate subtle metaphor, and induce the creating subject to produce a "flattened text" that is more easily visualized by AI and more in line with the expectation of data association: users generate images, algorithms recommend similar content, and instant feedback further incentivizes users to keep producing prompts and text forms that are easier to generate and easier to spread.

Literary reading is a typical activity of "delayed gratification"; the reader must follow the narrative rhythm and gain emotional resonance and intellectual enlightenment in the extension of time, which requires patience, concentration, and active imaginative investment. What "AI text-to-image" provides is "instant gratification" — input text and one instantly obtains a concrete result. AI-generated images are essentially spatialized and de-temporalized; they are frozen instants and slices stripped of forward and backward causality. When people grow accustomed to the instant feedback of obtaining visual answers with one click, their tolerance for time-consuming and effortful literary reading declines. Readers easily lose patience with those blank-leaving descriptions in literary works that require repeated savoring to appreciate their subtlety, and turn instead to crave or even demand rapid and certain visual transformation. The precious mental capacity, cultivated by literature, of steeping oneself in slow time and exploring in uncertainty is forcefully overridden by instant pleasure; the precious quality of literary experience — introspection through language and deep dialogue with self and world — is also lost in this process. Thus, literature gradually retreats from the central domain of meaning generation and exchange, and is downgraded by some platforms into a source of prompts for visual production. The standard for measuring literary value thereby changes as well: it no longer depends on intrinsic linguistic art, intellectual depth, and emotional tension, but is increasingly dominated by its ability to be transformed into eye-catching images. One possible consequence is a shift in society's overall "literary literacy": people's perception and appreciation of linguistic subtlety, narrative complexity, and emotional stratification are replaced by the pursuit of stunning spectacle, rhythmic pleasure, and direct emotional delivery.

Under the above orientation, the original impulse of literary creators — to encode with difficulty in language and to refine relentlessly in wording — tends toward decline. Literary creators begin to prefer deriving text backward from visual inspiration, rather than letting images gestate from within the text; narrative, instead of obeying the inner logic of language, turns to serve the creation of visual effects. The source of literary creation runs the risk of shifting from inner experience to an external database.

DOI: 10.54254/2753-7064/2026.39390

Literary style is the imprint of the author's unique worldview, emotional structure, and linguistic habit. Yet AI generation technology has a parametric and decontextualizing tendency, stripping style from the author so that it can be switched at will. If a user asks it to "paint in the style of Lu Xun," AI often only captures visual symbols and cannot reproduce Lu Xun's critical spirit. When style can be made concrete to the point of being applied at random, its profound connection with the ontological essence of life is weakened. Thus, the most precious personal linguistic pattern, unique rhythm, and ineffable "lingering charm" in the fabric of literary works are easily worn away under the power of the database.

4. The dissemination of simulacra and the erosion of cultural memory

Although "AI text-to-image" technology can break conventional logic and generate surreal, hybrid imagery, providing a new visual vocabulary for literary expression, this may stimulate writers and artists to break through the limits of their own imagination and develop entirely new metaphorical and symbolic systems, and can be regarded as an expansion of literary expressiveness. Yet, against the background of the rapid spread of convenient AI technology, the quicksand-like spectacle constituted by synchronic images is replacing the diachronic cultural mental ecology nourished by literature. Literature may be marginalized into a mere garnish in the visual feast, and may even hand over a large amount of collective cultural memory concerning its interpretation and inheritance to artificial intelligence.

When restoring literary scenes, the perspective method of painting and the reproductive technique of photography always maintain a referential relation with an external "real" world. Yet the images generated by "AI text-to-image" technology do not point directly to external reality, but point to image data that has already been collected, labeled, compressed, and recombined. This is precisely the "hyperreality" [5] foretold by Baudrillard — the model precedes the real and ultimately replaces it. The more decisive point is that platform incentives and popular expectation tend to reward those images that are clear, instant, recognizable, and transmissible, thereby forming an "averaged" visual default, causing the fragments of the spectacle to grow ever more homogenized. In this sense, the massive, rootless "simulacra" submerge us, yet do not necessarily truly enrich visual culture.

Because "a picture, unlike a sentence, is irrefutable... a picture does not advance an opinion, it implies no opposite to itself and no negation of itself, it obeys no rule of evidence or logic" [6], images possess powerful visual persuasiveness and communicability. When people widely use AI to illustrate literary classics, algorithmically generated images are quietly participating in and even dominating the construction of collective cultural memory. For example, AI-generated images of the Grand View Garden in *A Dream of Red Mansions* or the Trojan War in *Homer's epics* compete with the interpretations of painters across the ages and the space of verbal imagination, and are even easily recommended by platforms and shaped into high-visibility default images, creating a dangerous illusion of a sense of the real. When the "perfect" literary scenes generated by AI and conforming to popular expectation are everywhere, they compete with literary texts for position in the collective unconscious, and are even regarded by the public as a new "real." This is no different from using algorithms to render the cultural connotation and historical memory carried by literary works singular and superficial, challenging aesthetic autonomy and cultural inheritance.

If collective cultural memory partly shifts from textual understanding to AI-generated images, the inheritance and development of literature will face new risks. AI's visualizing behavior reinforces a latent logic: cultural memory that is difficult to visualize is at a disadvantage within the algorithmic visibility mechanism. This forces us to reconfirm: what is most precious about literature is precisely that part of mystery which resists complete visualization and resists being thoroughly said.

DOI: 10.54254/2753-7064/2026.39390

5. Conclusion

The fundamental contradiction highlighted by AI text-to-image lies in the conflict between the efficiency, clarity, and computability it pursues and the ambiguity, polysemy, and ineffability cherished by literary art. As Postman says, "If the users of a medium have understood its dangers, then that medium will not be overly dangerous" [7]. Although "AI text-to-image" technology may compress literature's meaning-space, it may also assist literature's cross-media expression; it may both reinforce visual centrism and stimulate literature's reconfirmation of its own irreplaceability. In the era of "AI text-to-image," where "spectacles" pile up and "simulacra" weigh heavy, this technological shift should not be rejected in a nostalgic manner. Instead, the flesh-and-blood imprint based on real life experience within literature should be defended, keeping consciousness complex, profound, and free.

The infinite interpretive space opened up by the metaphoricity of language, and the thickness of life borne by narrative time, among others, are the precious values of literature. "All genuine understanding originates in our not accepting what the world presents on its surface... language is a tool used to challenge, discuss, and question those surface things in our eyes." [7] By persisting in deep reading and complex writing, and by cherishing experiences difficult to visualize, the practice not only resists the discipline of algorithms, but also defends a spiritual reserve plot capable of complex thinking and subtle emotional experience.

The fate of literature is profoundly linked with the fate of the human subject: in an era when algorithms increasingly weave the net of perception, guarding the depth of literature bears on whether humans can continue to be complete human beings capable of deep thinking, delicate feeling, and free speech. How to preserve a spiritual life that can inhabit time, inhere in language, and remain forever open to possibility is the question posed to us by the fate of literature.

References

[1] Debord, G. (2006). *The society of the spectacle* (Z. Wang, Trans.). Nanjing University Press.
[2] Zhou, Q. C. (Ed.). (2015). *Foreign literary theory and comparative poetics* (Vol. 3). Intellectual Property Publishing House.
[3] Lakoff, G., & Johnson, M. (2015). *Metaphors we live by (W. Z. He, Trans.)*. Zhejiang University Press.
[4] Foucault, M. (1999). Discipline and punish (B. C. Liu & Y. Y. Yang, Trans.). SDX Joint Publishing Company.
[5] Baudrillard, J. (2025). *Simulacra and simulation* (R. Q. Wang, Trans.). Nanjing University Press.
[6] Postman, N. (2015). *The disappearance of childhood* (Y. T. Wu, Trans.). CITIC Press.
[7] Postman, N. (2015). *Amusing ourselves to death* (Y. Zhang, Trans.). CITIC Press.