Qeios

Research Article

Human Artistic Discernment as an Evaluative Criterion in the Age of Artificial Intelligence

Bolim Lee[1]

1. Institute for Cultural Arts Education Engineering, Republic of Korea

The rapid expansion of artificial intelligence has profoundly transformed contemporary practices of evaluation in academic and artistic domains. While AI systems have significantly enhanced the efficiency, scale, and consistency of assessment, they simultaneously raise fundamental questions regarding meaning, value, and responsibility dimensions that resist full automation. In this context, the present study revisits artistic discernmentas a core human capacity and examines its relevance as an evaluative criterion in the age of artificial intelligence. This paper conceptualizes artistic discernment through a dual framework. First, from a philosophical perspective, artistic discernment is understood as a distinctly human form of judgment grounded in contextual interpretation, value-based reasoning, and ethical reflection. Unlike algorithmic evaluation, which operates through pattern recognition and probabilistic inference, artistic discernment entails the capacity to situate works within broader cultural, historical, and moral contexts, thereby enabling judgments that are meaning-oriented rather than merely data-driven. Second, the study explores the methodological visibility of artistic discernment. Although traditionally regarded as subjective and intuitive, discernment is argued to exhibit recurring evaluative patterns that can be partially structured and examined through statistical approaches. Using simulated expert-rating data in the domain of music evaluation, the paper illustrates how consensus(Kendall's W), internal consistency(Cronbach's α), discriminative capacity(many-facet Rasch modeling), and weighting structures(AHP and entropy-based methods) can be employed to demonstrate the operational stability of discernment without reducing it to a purely quantitative construct. Importantly, this study does not propose artistic discernment as an alternative to artificial intelligence, nor does it seek to quantify human judgment exhaustively. Rather, it positions discernment as a complementary evaluative capacity that addresses dimensions of meaning, value, and ethical responsibility beyond the reach of automated systems. By articulating

artistic discernment as both philosophically grounded and methodologically approachable, this paper contributes to ongoing discussions on human-centered evaluation and offers a conceptual foundation for rethinking the role of human judgment in AI-mediated assessment environments.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

I. Introduction

The rapid advancement of artificial intelligence has reshaped evaluative practices across academic, artistic, and professional domains. Automated systems now assist in scoring, ranking, and screening processes that were previously reliant on human judgment. In many contexts, AI-based evaluation promises efficiency, scalability, and consistency. Yet, alongside these advantages, a fundamental concern persists: whether such systems are capable of assessing meaning, value, and responsibility dimensions that extend beyond formalized criteria and computational logic.

This tension is particularly evident in the evaluation of artistic and scholarly work. While AI can generate texts, images, and music with remarkable technical proficiency, the question of whether such outputs possess artistic or intellectual significance remains unresolved. Evaluation in these domains does not merely involve detecting patterns or measuring formal features; it requires judgment that is sensitive to context, intention, originality, and ethical implications.

Within this context, the concept of artistic discernment offers a productive point of departure. Traditionally discussed within arts education and aesthetic theory, artistic discernment refers to the capacity to perceive qualitative differences, interpret meaning within context, and make value-laden judgments that cannot be fully articulated through rules or metrics. Rather than being a residual or purely subjective faculty, discernment has historically functioned as a central evaluative competence in artistic, educational, and cultural practices[1][2][3].

However, contemporary discussions often frame artistic discernment as incompatible with methodological rigor, positioning it in opposition to data-driven or algorithmic evaluation. This dichotomy risks marginalizing human judgment precisely at a moment when its role requires renewed articulation. The present study challenges this opposition by proposing that artistic discernment can be understood through a dual framework: as a philosophically grounded human capacity and as a form of judgment that exhibits discernible patterns capable of partial methodological articulation.

Accordingly, this paper pursues two interrelated aims. First, it examines artistic discernment as a philosophically distinctive mode of judgment characterized by contextual interpretation, value–based reasoning, and ethical reflection. Second, the study explores the methodological visibility of discernment by illustrating how its outcomes can be examined through statistical tools without collapsing its qualitative nature into purely quantitative terms.

To this end, the paper employs simulated expert-rating data in music evaluation to demonstrate how consensus, consistency, and discriminative capacity may serve as indicators of discernment in practice, drawing upon established approaches in measurement theory and reliability analysis[4][5][6][7].

By integrating philosophical analysis with methodological illustration, this study positions artistic discernment as a complementary evaluative criterion in AI-mediated environments. Rather than opposing artificial intelligence, the paper argues for a reconfiguration of evaluation in which human judgment retains responsibility for meaning, value, and ethical orientation. In doing so, it contributes to ongoing debates on human-centered evaluation and reassesses the role of human discernment in the age of artificial intelligence[8][9][10]. Educational policy frameworks have also emphasized interpretive and reflective competencies as essential human capacities in complex societies[11].

II. Theoretical Background

2.1. Artistic Discernment as a Human Mode of Judgment

Artistic discernment has long been discussed within aesthetic theory and arts education as a distinctive mode of human judgment. Rather than referring to subjective taste or individual preference, discernment denotes the capacity to perceive qualitative differences, interpret meaning within context, and make value-oriented evaluations. In this sense, discernment functions as a form of judgment that integrates perception, interpretation, and evaluation into a coherent act.

Eisner conceptualized artistic discernment as an essential educational capacity through which individuals learn to notice subtle distinctions, recognize significance, and make informed judgments that cannot be reduced to standardized criteria[3]. From this perspective, discernment is neither arbitrary nor purely intuitive; it is cultivated through experience, practice, and engagement with cultural forms. Importantly, such judgment resists full formalization, not because it lacks structure, but because its structure is inseparable from contextual understanding.

This conception aligns with broader philosophical accounts of human judgment. Dewey emphasized that evaluative judgment in art emerges from experience, where meaning is constructed through interaction rather than extracted through rule-based analysis[12]. Similarly, Polanyi's notion of tacit knowledge highlights the dimensions of knowing that operate beyond explicit articulation, yet remain central to competent judgment[13]. Artistic discernment, therefore, can be understood as a synthesis of experiential knowledge, interpretive awareness, and evaluative responsibility.

This view is also compatible with symbolic theories of art, in which meaning is understood as arising from systems of representation rather than purely perceptual features[14].

2.2. The Limits of Algorithmic Evaluation

The growing application of artificial intelligence in evaluative contexts has foregrounded questions concerning the nature and limits of algorithmic judgment. AI systems excel at identifying patterns, optimizing measurable criteria, and producing consistent outputs across large datasets. In domains where evaluation can be fully specified through formal rules or quantifiable indicators, such systems offer undeniable advantages.

However, evaluation in artistic and scholarly contexts involves dimensions that exceed computational description. Meaning, originality, ethical implication, and contextual relevance are not properties that can be exhaustively inferred from data patterns alone. While AI can model correlations among features, it does not possess an internal standpoint from which significance is interpreted or values are prioritized. As a result, algorithmic evaluation remains structurally constrained to what can be operationalized in advance.

This limitation does not imply that AI evaluation is inherently flawed, but rather that it operates within a different epistemic register from human judgment. Algorithmic systems assess according to predefined parameters, whereas artistic discernment involves the capacity to question the adequacy of those very parameters. In this respect, discernment functions not as a competing technology, but as a meta-evaluative capacity that reflects on the grounds of evaluation itself[8].

2.3. Ethical and Contextual Dimensions of Discernment

A further dimension that distinguishes artistic discernment from algorithmic evaluation lies in its ethical orientation. Judgment in artistic and academic contexts is not value-neutral; it entails responsibility for the consequences of evaluation, including recognition, exclusion, and legitimation.

Discernment therefore involves an awareness of how judgments operate within broader social, cultural, and institutional contexts.

Nussbaum's work on ethical judgment underscores the role of practical reasoning in situations where general rules prove insufficient[15][16]. Similarly, Biesta has argued that an overreliance on measurement risks displacing questions of purpose, responsibility, and educational meaning[17][18]. Within this framework, artistic discernment can be understood as a form of judgment that remains responsive to singularity and context, rather than seeking closure through abstraction.

In the age of artificial intelligence, these ethical dimensions acquire renewed significance. As automated evaluation systems increasingly mediate access to cultural and academic recognition, the question is not merely how accurately systems perform, but how evaluative authority is distributed and justified. Artistic discernment, in this sense, serves as a safeguard against the uncritical delegation of judgment to systems that cannot assume responsibility for their outcomes.

2.4. Toward a Dual Framework of Discernment

Taken together, these perspectives suggest that artistic discernment occupies a distinctive position within contemporary evaluation debates. On the one hand, it represents a philosophically grounded human capacity rooted in experience, interpretation, and ethical reflection. On the other hand, its repeated manifestation in evaluative practice suggests that discernment is not wholly idiosyncratic, but exhibits patterns that invite methodological examination.

Rather than treating discernment as either an ineffable intuition or a fully quantifiable skill, the present study approaches it as a mode of judgment that is conceptually irreducible yet methodologically approachable. Such an approach avoids the false dichotomy between subjectivity and objectivity, allowing discernment to be analyzed without undermining its qualitative integrity.

By situating artistic discernment within this dual framework, the study provides a theoretical foundation for the methodological illustrations presented in the subsequent section and prepares the ground for examining how human judgment can be articulated, rather than replaced, within AI-mediated evaluative environments[19][20].

III. The Dual Framework of Artistic Discernment: Philosophical Singularity and Methodological Visualization

3.1. Introduction: At the Boundary Between Humans and AI

The rapid expansion of artificial intelligence in the twenty-first century has fundamentally unsettled established frameworks for academic and artistic evaluation. Texts, images, and music can now be generated with ease; however, whether such outputs possess genuine scholarly or artistic significance remains a separate and unresolved question. At this juncture, artistic discernment is recalled not as a matter of subjective taste, but as a criterion through which humanity defines and evaluates itself.

This chapter examines artistic discernment along two interrelated axes. First, at the level of philosophical singularity, discernment functions as a final stronghold of human judgment. Second, through methodological visualization, discernment does not remain confined to pure subjectivity but can be partially revealed as a pattern of judgment open to analytical scrutiny.

3.2. The Philosophical Singularity of Artistic Discernment

3.2.1. Contextual Judgment

While AI systems are capable of enumerating facts, they are unable to interpret meaning. Artistic discernment enables the evaluator to grasp the context in which a work or study is produced and to perceive the direction of value embedded within that context. This capacity constitutes a form of judgment intrinsic to human cognition and is fundamentally distinct from data–driven mechanical outputs[12][13].

3.2.2. Value Judgment

The authenticity, originality, and scholarly or artistic contribution of a paper or artwork cannot be readily reduced to numerical indicators. Artistic discernment operates by identifying qualitative distinctions and differentiating what is meaningful from what is not. In this respect, discernment serves as a critical faculty that resists simplistic quantification[3].

3.2.3. Ethical Reflection

Outputs generated by AI inherently carry the risk of conflating fact and fabrication. Artistic discernment extends beyond evaluation in a narrow sense, enabling judgments that take into account social responsibility and ethical legitimacy. From this perspective, discernment constitutes an inviolable domain of human judgment[15][17].

3.3. Methodological Visualization and Statistical Application of Artistic Discernment

Artistic discernment has traditionally been understood as an intuitive and subjective form of insight. However, examination of evaluative outcomes reveals recurring patterns suggesting that discernment functions as an indicator of social reliability rather than merely individual preference. In particular, consensus, consistency, and discrimination emerge as key indicators of evaluative judgment, drawing upon established statistical reliability and measurement frameworks[4][5][6][7].

These statistical results do not quantify judgment itself but demonstrate the stability of evaluative agreement. At a different level, multi-criteria decision approaches and fuzzy evaluation theories provide conceptual models through which structured judgment in complex evaluation environments can be represented[21][22].

To illustrate this point, simulated data are presented in which five experts evaluate six musical works using a five-point scale.

Work/ Expert	Expert A	Expert B	Expert C	Expert D	Expert E
Work1	4	5	4	5	4
Work 2	3	3	2	3	3
Work 3	5	4	5	5	5
Work 4	2	2	3	2	2
Work 5	4	4	4	3	4
Work 6	1	2	1	1	1

Table 1. Example of Simulated Data (Expert Ratings of Musical Works)

When analyzed using Kendall's W, Cronbach's α, and a many-facet Rasch model, the following results emerge.

First, Kendall's W yields a value of 0.78, indicating a high level of consensus among experts. This suggests that artistic discernment in music evaluation does not remain at the level of individual preference but can operate as a shared evaluative criterion to a certain extent.

Second, Cronbach's α is calculated as 0.84, confirming that consistency is maintained even when evaluative criteria are applied repeatedly. This finding supports the claim that discernment is not a momentary intuition but a stable and reproducible judgmental capacity.

Third, the application of the Rasch model reveals that Work 3 occupies the upper end of the evaluative scale (predominantly receiving scores of 4–5), whereas Work 6 forms the lower extreme (predominantly receiving scores of 1–2). This distribution demonstrates that artistic discernment effectively discriminates between qualitative differences among works.

3.3.1. Consensus

The degree of agreement among expert evaluations can be measured using Kendall's coefficient of concordance(W).

$$W = \frac{12 \sum_{j=1}^{n} \overline{R_j} \left(R_j - \overline{R} \right)^2}{m^2 \left(n^3 - n \right)}$$

where m denotes the number of raters, n the number of evaluated items, and R_j the sum of ranks assigned to the j-th item. Applying this measure to the simulated data in which five experts evaluated six musical works yields Kendall's W = 0.78. This result structurally demonstrates that artistic discernment can function as a socially shareable evaluative criterion rather than remaining at the level of idiosyncratic individual preference.

3.3.2. Consistency

The stability of evaluative judgments across items can be examined using Cronbach's alpha (α).

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i-1}^{k} \sigma_i^2}{\sigma_T^2} \right)$$

Here, k represents the number of items, $\sigma_i{\wedge}2$ the variance of each item, and $\sigma\_T{\wedge}2$ the variance of the total score, where $T = \sum_{(i=1)}{\wedge}k X\_i$. When applied to the simulated data, Cronbach's $\alpha = 0.84$, indicating a high level of internal consistency among the evaluation items. A high alpha value suggests that artistic discernment is not a momentary intuition but a reproducible and reliable judgmental capacity. In other words, the tendency of evaluators to produce similar results across repeated assessments supports the stability of discernment as a human ability.

3.3.3. Discrimination

The ability of artistic discernment to distinguish qualitative differences among evaluated objects can be analyzed using the Many-Facet Rasch Model (MFRM).

$$\log \frac{P_{nij}(x)}{P_{nij}(x-1)} = B_i - D_j - C_n - F_x$$

In this equation, B_i represents the ability of the evaluated object, D_j the difficulty of the item, C_n the severity of the rater, and F_x the scale step parameter. Within the simulated data, Work 3 received the highest scores (predominantly 4–5), while Work 6 received the lowest scores (predominantly 1–2), thereby forming the two extremes of the evaluative scale. The Rasch analysis demonstrates that artistic discernment does not assign uniformly similar evaluations but instead reveals concrete qualitative differences among works. Securing such discrimination indicates that discernment plays a substantive role in artistic value judgment.

3.3.4. Weight Structuring

The relative importance of the components underlying artistic discernment can be derived by combining Analytic Hierarchy Process (AHP) weights with entropy-based weights.

- AHP weighting:

$$A_\omega = \lambda_{max\omega} \, , \; \sum_j \omega_j^{AHP} = 1$$

- Consistency ratio:

$$CR = \frac{\lambda_{\max} - n}{(n-1)} \; CR = \frac{CI}{RI} = \frac{\lambda_{\max} - n}{(n-1)RI}$$

* CR ≤ 0.1 indicates acceptable consistency.

- Entropy weighting:

$$e_j = -k \sum_i p_{ij} \, ln p_{ij} \quad d_j = 1 \, - \, e_j \quad \omega_j^{ENT} = \frac{d_j}{\sum d_j}$$

- Combined weighting:

$$\omega_j \; = \; \lambda \omega_j^{AHP} + (1 - \lambda) \omega_j^{ENT}$$

In the simulated application, the resulting weights were as follows: creativity (0.35), completeness (0.30), emotional expressiveness (0.20), and contextual appropriateness (0.15). These results illustrate that artistic discernment operates in a balanced manner between intuitive expert judgment (subjectivity) and data-based differentiation (objectivity). In this sense, discernment embodies a structural property that simultaneously accommodates both subjective and objective dimensions.

3.3.5. Reliability and Validity

A Generalizability Theory (G-theory) analysis yielded a generalizability coefficient of G = 0.76, indicating adequate reproducibility across raters and items. In addition, confirmatory factor analysis (CFA) supported the three-factor structure of consensus, consistency, and discrimination, with fit indices of CFI = 0.92 and RMSEA = 0.07, both within acceptable ranges. These findings confirm that the structure of artistic discernment satisfies established criteria for reliability and validity.

3.3.6. Synthesis

Taken together, these analyses demonstrate that artistic discernment retains its philosophical singularity while also allowing partial visualization through statistical tools. Although illustrated here through musical evaluation, this approach can be extended to other artistic domains such as visual arts, literature, and film, and further to broader contexts of judgment including academic evaluation and policy assessment. Ultimately, artistic discernment may be understood as a universal judgment system through which humans identify meaning and value within complex contexts.

IV. Conclusion: Reframing Human Evaluation in the Age of Artificial Intelligence

This study has examined how artistic discernment, as a distinctly human capacity for judgment, may be understood and repositioned within contemporary academic and artistic evaluation environments shaped by artificial intelligence. While rapid advances in AI technologies have dramatically enhanced the

efficiency and scalability of evaluation systems, they have simultaneously introduced the risk of marginalizing human-centered dimensions of judgment such as meaning interpretation, value orientation, and ethical responsibility.

In response to this concern, the present study reconceptualizes artistic discernment not as a matter of subjective taste or intuition, but as a central criterion for human evaluation in the age of artificial intelligence. At the level of philosophical singularity, discernment encompasses contextual judgment, value judgment, and ethical reflection grounded in experience and interpretation[12][13]. These characteristics highlight the fundamental limitation of artificial intelligence: while AI systems can process information and detect patterns, they cannot independently determine meaning or establish value hierarchies[8].

The study further proposed a methodological pathway through which artistic discernment may be partially visualized without reducing it to purely quantitative measurement. Structural indicators such as consensus, consistency, and discrimination demonstrate that discernment exhibits recurring evaluative patterns observable in practice. Analytical approaches in reliability and measurement theory support the interpretation that judgment can be structurally described while retaining qualitative integrity[4][5][6][7].

Importantly, this study does not present artistic discernment as a substitute for AI-based evaluation. Rather, discernment is positioned as a human-centered evaluative capacity that complements artificial intelligence by addressing its limitations and safeguarding ethical and interpretive dimensions of evaluation. This perspective challenges the assumption that increased automation necessarily entails the disappearance of human judgment and instead suggests a cooperative structure of evaluation between algorithmic processing and human interpretation[9][10].

Although musical evaluation served as the primary illustrative case, the proposed framework may be extended to other artistic and scholarly domains and to broader evaluative contexts including academic review and professional assessment. Such a view aligns with contemporary educational discussions emphasizing interpretive understanding and reflective judgment[19][20].

V. Epilogue: Leaving Judgment Behind

Artificial intelligence has expanded the speed and scale of evaluation to an unprecedented degree. Yet the questions of what ought to be evaluated and what meaning such evaluation ultimately carries remain irreducibly human concerns. The return to the concept of artistic discernment in this study does not

represent nostalgia, but rather a response to a growing void of judgment in technologically mediated evaluation environments.

Artistic discernment cannot be reduced to a complete set of rules, nor does it align all judgments under a single uniform standard. Instead, it involves interpreting contexts, distinguishing qualitative differences, and assuming responsibility for evaluative decisions. These characteristics correspond to forms of practical reasoning and ethical responsibility emphasized within humanistic and educational philosophy[15][16][17][18].

The aim of this study, therefore, is not to oppose artificial intelligence but to clarify the domain that should not be delegated entirely to automated systems. As evaluation processes become increasingly algorithmic, it becomes equally necessary to determine what must remain within human responsibility. Artistic discernment functions as a minimal but essential human judgmental capacity that preserves meaning, value, and accountability in an era of technological assessment[8].

Ultimately, to discuss artistic discernment is to reconsider the criteria by which significance is recognized and judgments are justified. As long as evaluation concerns meaning rather than mere calculation, human judgment will continue to occupy a central place in intellectual and cultural life.

References

1. ^Eisner EW (1972). *Educating Artistic Vision. Macmillan.*

2. ^Eisner EW (1985). *The Educational Imagination: On the Design and Evaluation of School Programs. 2nd ed. Macmillan.*

3. a, b, cEisner EW (2002). *The Arts and the Creation of Mind. Yale University Press.*

4. a, b, cKendall MG, Smith BB (1939). *"The Problem of M Rankings." Ann Math Stat.* **10***(3):275–287. doi:10.1214/ aoms/1177732186.*

5. a, b, cCronbach LJ (1951). *"Coefficient Alpha and the Internal Structure of Tests." Psychometrika.* **16***(3):297–334. doi:10.1007/BF02310555.*

6. a, b, cLinacre JM (1989). *"Many-Facet Rasch Measurement." University of Chicago.*

7. a, b, cLinacre JM (2014). *"Winsteps® Rasch Measurement Computer Program." Winsteps.com.*

8. a, b, c, dFloridi L (2014). *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality. Oxford University Press.*

9. a, b Floridi L, Cowls J, Beltrametti M, et al. (2018). "AI4People—An Ethical Framework for a Good AI Society." *Minds Mach.* **28**(4):689–707. doi:10.1007/s11023-018-9482-5.

10. a, b UNESCO (2021). *Reimagining Our Futures Together: A New Social Contract for Education. UNESCO Publishing.*

11. ^ OECD (2019). *OECD Learning Compass 2030. OECD Publishing.*

12. a, b, c Dewey J (1934). *Art as Experience. Minton, Balch & Company.*

13. a, b, c Polanyi M (1966). *The Tacit Dimension. University of Chicago Press.*

14. ^ Goodman N (1976). *Languages of Art: An Approach to a Theory of Symbols. 2nd ed. Hackett Publishing.*

15. a, b, c Nussbaum MC (1990). *Love's Knowledge: Essays on Philosophy and Literature. Oxford University Press.*

16. a, b Nussbaum MC (2011). *Creating Capabilities: The Human Development Approach. Harvard University Press.*

17. a, b, c Biesta G (2010). *Good Education in an Age of Measurement: Ethics, Politics, Democracy. Paradigm Publishers.*

18. a, b Biesta G (2015). *The Beautiful Risk of Education. Routledge.*

19. a, b McTighe J, Wiggins G (2005). *Understanding by Design. 2nd ed. ASCD.*

20. a, b Wiggins G (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance. Jossey-Bass.*

21. ^ Saaty TL (1980). *The Analytic Hierarchy Process. McGraw-Hill.*

22. ^ Zadeh LA (1965). "Fuzzy Sets." *Inf Control.* **8**(3):338–353. doi:10.1016/S0019-9958(65)90241-X.

Declarations