2026 Volume: 02 Issue: 06 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-36

FROM AESTHETIC JUDGMENT TO STRATEGIC FORESIGHT: A BAYESIAN FRAMEWORK FOR INTEGRATING ART CRITICISM INTO BUSINESS MANAGEMENT EDUCATION

***Dr. Chitranka K.**

Assistant Professor, NSB WBS.

Article Received: 8 May 2026
Article Revised: 28 May 2026
Published on: 18 June 2026

***Corresponding Author: Dr. Chitranka K.**
Assistant Professor, NSB WBS.

DOI: https://doi-doi.org/101555/ijrpa.7281

ABSTRACT

We present a computational framework that integrates art criticism into business management education by modeling aesthetic judgment as a dynamic Bayesian network. The conventional reliance on subjective exercises such as theater improvisation or design thinking workshops often fails to produce transferable, quantifiable decision-making skills. To address this gap, we propose a system that operationalizes the sequential phases of aesthetic evaluation—perception, contextual analysis, interpretative synthesis, and evaluative conclusion—as latent states in a temporal probabilistic model. A pre-trained multimodal transformer encodes visual and textual business artifacts into aesthetic descriptor vectors, which then drive the network's state transitions. A key novelty is the inclusion of an epistemic expectation integrator that tracks the learner's shifting confidence through Kullback–Leibler divergence between successive posterior distributions, thereby mirroring the critical re-evaluation stage central to professional art criticism. The system maps the evolving aesthetic appraisal onto a binary leadership competency indicator, which is then compared against conventional market benchmarks through a divergence score. A pedagogical feedback generator subsequently produces structured critique prompts using a gradient-boosted tree model and a fine-tuned large language model, enabling iterative refinement of the learner's judgment. We further construct a Standardized Taxonomy of Aesthetic-Leadership Indicators from accumulated session data, providing a data-driven reference for artistic cognition in business pedagogy. Our framework transforms an intuitive, tacit process into a repeatable, quantifiable tool for strategic decision-making under high ambiguity. This approach thus bridges the gap between aesthetic sensitivity and business leadership, offering both a rigorous methodology for

educational assessment and a novel pathway for identifying creative agility in future managers.

INTRODUCTION

Contemporary business management education faces significant pressure to cultivate leaders capable of navigating increasingly volatile and ambiguous markets. Traditional curricula, heavily reliant on quantitative analysis and rational decision-making models, have been criticized for their inability to prepare students for the nuanced, intuitive judgments required in complex strategic scenarios (Grey, 2004). This critique has spurred interest in arts-based learning (ABL), a pedagogical movement that incorporates artistic practices—such as theater improvisation, visual storytelling, or musical composition—into management training to enhance creativity, empathy, and pattern recognition (Nissley, 2010). While numerous studies have qualitatively documented the benefits of ABL for fostering soft skills like creative confidence and design thinking aptitude (Cuenca et al., 2014), a persistent challenge remains: how to transform the inherently subjective, tacit nature of aesthetic judgment into a transferable and quantifiable competency for strategic decision-making.

The core problem lies in the operationalization of aesthetic perception—a process that professional art critics perform through distinct, sequential phases of observation, interpretation, and evaluation. In management education, this process is often reduced to unstructured exercises, such as attending a gallery visit or engaging in free-form roleplay, without a rigorous mechanism to capture or assess how students translate aesthetic cues into actionable business insights (Katz-Buonincontro, 2015). Consequently, the link between *artistic sensitivity* and *leadership effectiveness* remains largely anecdotal, lacking the empirical grounding necessary for widespread curricular adoption. We require a framework that not only replicates the structured pathways of professional art criticism but also generates data that can be correlated with traditional business performance indicators.

To address this gap, we propose a computational framework that models the sequential phases of aesthetic judgment as a dynamic Bayesian network (DBN). DBNs are well-suited for this task because they can represent probabilistic dependencies between hidden cognitive states (e.g., "emotional resonance" or "compositional tension") and observable actions over time, updating beliefs as new information is processed (Murphy, 2002). Our system takes as input qualitative visual and textual cues derived from complex market scenarios—such as a competitor's new branding campaign or a crisis memo. A pre-trained multimodal transformer encodes these artifacts into aesthetic descriptor vectors, which serve as observations for the

DBN. The network tracks the learner's evolving cognitive state as they progress through four stages: perception, contextual analysis, interpretative synthesis, and evaluative conclusion, mirroring the rigor of professional art criticism.

A central innovation is the *epistemic expectation integrator*, a module that continuously monitors the divergence between successive posterior probability distributions using the Kullback–Leibler (KL) divergence. This metric quantifies how the learner's confidence shifts during the evaluation, thereby operationalizing the critical re-evaluation stage that separates novice from expert judgment. The system then maps the final posterior state onto a binary leadership competency indicator, such as "strategic foresight" or "ambiguity tolerance," and compares this against conventional market benchmarks (e.g., net present value calculations) through a divergence score. This score, combined with a gradient-boosted tree model that identifies key diagnostic cues, feeds into a pedagogical feedback generator that uses a fine-tuned large language model to produce structured, Socratic critique prompts (García, 2024). The objective is not to replace human intuition but to make its process visible, debatable, and systematically improvable.

Our framework contributes a novel bridge between the *qualitative* tradition of arts-based learning and the *quantitative* demands of business education. First, it provides a systematic methodology for transforming subjective aesthetic impressions into quantifiable, tractable metrics for educational assessment. Second, it introduces a data-driven approach to constructing a Standardized Taxonomy of Aesthetic-Leadership Indicators (STALI) from accumulated session data, offering a shared vocabulary for educators to discuss artistic cognition in business contexts. Third, by embedding structured comparative feedback within an iterative pedagogical loop, the system directly aligns with social constructivist theories of learning, where knowledge is co-constructed through peer critique and reflection (Dyllick, 2015). This approach moves beyond 'decorative integration'—where art is merely a motivational tool—towards 'relevant learning', in which artistic practice directly enhances managerial foresight.

The remainder of this paper is organized as follows. Section 2 reviews related work in arts-based management education, computational aesthetics, and dynamic belief modeling. Section 3 establishes the theoretical foundations, detailing the connection between the sequential phases of professional art criticism and Bayesian inference under uncertainty. Section 4 presents the architecture of our dynamic Bayesian framework, including the epistemic expectation integrator and the pedagogical feedback generator. Section 5 describes our experimental design, evaluation metrics, and preliminary findings from a pilot study with MBA students. Section 6 discusses the implications of our results, the limitations of the current approach, and potential avenues for future research. Finally, Section 7 concludes by summarizing how our method bridges the gap between artistic intuition and strategic leadership.

Related Work

The integration of arts into management education has a considerable history, yet its operationalization for strategic decision-making remains nascent. This section reviews three relevant streams of literature: arts-based learning in business schools, computational approaches to aesthetic judgment, and probabilistic models of sequential decision-making under uncertainty.

Arts-Based Learning and Aesthetic Leadership in Management Education

Management education has increasingly incorporated artistic practices to foster creativity, empathy, and pattern recognition—skills deemed essential for navigating high-ambiguity environments (Nissley, 2010). These methods, collectively termed arts-based learning (ABL), range from theater improvisation exercises that build collaborative intuition to visual art workshops that enhance perceptual acuity. However, a recurring critique in the literature is that such interventions often remain "decorative" rather than "relevant," lacking a rigorous mechanism to measure their transfer to business competencies (Katz-Buonincontro, 2015). For example, studies report that students in design thinking courses develop creative confidence, but the specific cognitive processes involved—such as how they move from initial perception to evaluative conclusion—are typically left unexamined (Cuenca et al., 2014). This gap is particularly salient in the growing field of aesthetic leadership, which posits that a leader's capacity for refined aesthetic judgment directly influences organizational integrity and clinical decision-making in professional settings (Bashandy et al., 2024). Some researchers have proposed that aesthetic judgment is best understood as a form of decision-making that unfolds over temporally distinct phases, involving perception, contextual analysis, interpretative synthesis, and evaluative conclusion (Nojo, 2026). However, existing pedagogical frameworks for aesthetic leadership rarely operationalize these phases in a manner that can be systematically taught or assessed within business curricula (Quijada et al., 2024). Our work directly addresses this limitation by modeling these sequential phases as a dynamic Bayesian network, thereby creating a computational scaffold for aesthetic reasoning that is both teachable and measurable.

Computational Models of Aesthetic Perception and Judgment

Parallel to developments in ABL, computational aesthetics has made significant strides in modeling human evaluative processes using machine learning. Pre-trained multimodal transformer architectures, such as CLIP (Contrastive Language-Image Pre-training), have demonstrated remarkable ability to encode visual and textual artifacts into shared embedding spaces that capture abstract features like “visual tension” or “narrative coherence” (Radford et al., 2021). These embeddings have been successfully deployed for tasks ranging from predicting aesthetic ratings of artwork to generating stylistically consistent product designs. However, such models typically operate as static classifiers, assigning a single score or label to an artifact without modeling the temporal evolution of an observer’s judgment. In the context of education, practitioners have called for systems that can capture the iterative, re-evaluative nature of expert judgment, where initial impressions are systematically refined through successive phases of analysis (Nojo, 2026). Our framework extends computational aesthetics by embedding a multimodal encoder within a dynamic Bayesian network, allowing the same visual cue to be re-interpreted at different phases of the judgment process, with each phase updating the latent appraisal state. This temporal modeling is a key departure from static aesthetic classifiers and aligns more closely with empirical evidence on how professional critics develop their evaluations.

Dynamic Bayesian Networks for Modeling Sequential Decision-Making Under Uncertainty

Dynamic Bayesian networks (DBNs) are a well-established formalism for modeling probabilistic systems that evolve over time, particularly when the underlying state is hidden and only observable through noisy measurements (Murphy, 2002). They have been applied extensively in fields such as robotics, speech recognition, and medical diagnosis to infer latent cognitive or physical states from sequential observations. In the domain of decision-making under uncertainty, researchers have used DBNs to model how individuals update their beliefs when confronted with new evidence, often incorporating parameters for confidence, surprise, or cognitive load (Frank, 2013). However, the application of DBNs to *aesthetic* judgment—a domain traditionally seen as subjective and rule-averse—is relatively unexplored. Some recent work has used Bayesian approaches to model aesthetic preferences as probabilistic distributions over feature space, but these models typically assume a single-shot evaluation rather than a phased, sequential process (Iigaya et al., 2021). Our contribution integrates these two lines of research by constructing a DBN whose latent states correspond to the discrete phases of art criticism, and whose transition dynamics are modulated by an epistemic expectation integrator that computes the Kullback-Leibler divergence between successive posteriors. This integration naturally captures the "cognitive tension" that arises when new aesthetic cues challenge prior assessments, a phenomenon well-documented in the psychology of aesthetic evaluation (Nojo, 2026).

Key Novelty Over Existing Approaches

While existing work has separately advanced arts-based learning (Nissley, 2010), computational aesthetics (Radford et al., 2021), and probabilistic decision modeling (Murphy, 2002), no framework has yet unified these streams to create an *actionable pedagogical tool* for training aesthetic judgment in business leaders. Unlike conventional ABL interventions, which rely on subjective introspection and unstructured debriefing (Katz-Buonincontro, 2015), our system provides quantifiable, traceable outputs that can be correlated with traditional business KPIs. Unlike static aesthetic classifiers (Iigaya et al., 2021), it captures the temporal, re-evaluative nature of expert judgment. And unlike general DBN-based decision models (Frank, 2013), it operates on *aesthetic descriptors* rather than raw market data, thereby bridging the gap between artistic sensitivity and strategic foresight. The central innovation is the epistemic expectation integrator, which operationalizes the critical re-evaluation phase of professional criticism as a measurable divergence metric, and the pedagogical feedback generator, which transforms this metric into structured learning prompts. This combination transforms aesthetic intuition from an opaque talent into a trainable, data-informed competency.

Theoretical Foundations: Aesthetic Judgment and Bayesian Inference in Decision-Making

To build a rigorous computational framework for arts-based management education, we must first establish the theoretical equivalences between the sequential phases of professional art criticism and the probabilistic mechanics of Bayesian inference. This section lays the epistemological and mathematical groundwork by examining three interconnected pillars: the nature of tacit knowledge in decision-making, the dynamic temporal logic of evaluation under uncertainty, and the geometric representation of subjective qualities.

The Nature of Tacit Knowledge and Latent Cognitive States in Strategic Decision-Making

Strategic decisions in business environments rarely rely on complete, unambiguous information. Instead, they frequently draw upon tacit knowledge—the unarticulated, intuitive understanding that experts accumulate through experience but cannot fully explicate (Polanyi, 2009). This epistemological challenge is particularly acute in aesthetic judgment, where an experienced art critic might identify a “lack of compositional harmony” or an “unresolved narrative tension” without being able to specify the algorithmic rules underlying such assessments. In management education, this tacit dimension manifests when a seasoned leader senses that a new market entrant’s branding is “inauthentic” or that a crisis communication feels “disingenuous,” yet struggles to articulate the precise cues that triggered this intuition.

Linear regression models, decision trees, and rule-based expert systems—the traditional tools of business quantitative analysis—fail to capture these nuanced internal transitions. They treat decisions as functions of explicit, pre-defined features, ignoring the iterative, self-referential process through which human judges refine their own initial impressions. Professional art criticism, by contrast, unfolds through distinct phases: a rapid perceptual scan, a contextual anchoring, a period of interpretative synthesis, and a final evaluative conclusion (Nojo, 2026). Each phase modifies the cognitive representation of the artifact in ways that are not directly observable. Hence, we frame aesthetic judgment as a sequence of hidden mental states—latent cognitive states—that drive observable decision behaviors (such as verbal critiques or strategic choices).

Probabilistic Temporal Dynamics in Uncertain Environments

Given that strategic decision-making under high ambiguity involves the sequential integration of ambiguous cues over time, static models that assume a one-shot evaluation are fundamentally inadequate. Imagine a business leader evaluating a competitor’s new marketing campaign: initial impressions based on color palette and typography may shift dramatically after contextual information about the brand’s recent history is considered, only to be revised again when market performance data becomes available. This temporal dynamic mirrors the re-evaluative process central to art criticism, where initial aesthetic pleasure is frequently revised through interpretive reasoning (Pelowski & Akiba, 2011).

Dynamic Bayesian networks are specifically designed to model such scenarios. They maintain a probability distribution over a set of hidden states (e.g., “aesthetic resonance,” “narrative coherence,” “contextual relevance”) and update this distribution as new observations arrive over discrete time steps (Murphy, 2002). The transition between states is governed by a conditional probability matrix [illegible], capturing how one phase of judgment naturally flows into the next. Unlike recurrent neural networks, which approximate this process through opaque weight matrices, DBNs offer interpretable conditional dependencies that can be inspected, modified, and validated by domain experts. We therefore hypothesize that the four-phase structure of aesthetic judgment—perception, contextual analysis, interpretative synthesis, evaluative conclusion—can be fruitfully encoded as a four-state DBN, where each phase corresponds to a distinct latent state with its own emission probabilities for observable behaviors.

Geometric Representation of Semantic and Aesthetic Concepts

A central challenge in operationalizing aesthetic judgment is the representation of qualitative, subjective concepts—such as “visual tension,” “thematic fragmentation,” or “historical resonance”—in a form that is computable yet retains semantic meaning. We adopt the hypothesis that these qualitative descriptors can be mapped to continuous vectors in a high-dimensional embedding space, where geometric relationships between vectors correspond to semantic similarities (Radford et al., 2021). A pre-trained multimodal transformer, such as CLIP, provides precisely this capability: it maps both images and their text descriptions into a shared embedding space, such that the vector for an image of a stormy sea closely aligns with the vector for the text phrase “chaotic, unresolved tension.”

Under this geometric representation, the sequential phases of aesthetic judgment can be reinterpreted as a trajectory through the embedding space. The initial perceptual phase places the learner’s cognitive state near a region corresponding to surface-level features (e.g., “bright colors,” “symmetrical composition”). The contextual analysis phase shifts this state towards a region encoding historical or cultural references. The interpretative synthesis phase moves it into a region representing abstract meaning (e.g., “metaphor for corporate upheaval”). And the evaluative conclusion phase maps it onto a region correlated with a decision outcome—such as “recommend further investigation” or “dismiss as irrelevant.” This trajectory provides a geometric trace of the cognitive process, which we can model as a stochastic path within the embedding space. The Bayesian update rule then governs how each new visual or textual observation shifts the estimated position of the learner’s hidden state within this space. Because the embedding space is continuous and high-dimensional, it naturally accommodates the nuance and ambiguity of aesthetic concepts, unlike discrete categorical systems that force all evaluations into a few rigid boxes.

Connecting these three theoretical pillars, we observe a coherent narrative: tacit knowledge manifests as unobserved latent states; these states evolve over time in a probabilistic manner that DBNs can capture; and the content of these states can be meaningfully projected onto geometric representations of aesthetic concepts. This unified theoretical foundation motivates the architectural design we present in the following section.

A Dynamic Bayesian Framework for Operationalizing Art-Critical Phases in Management Training

Building upon the theoretical foundations established above, we now present the technical architecture of our proposed system. The framework is designed to emulate the sequential phases of professional art criticism—perception, contextual analysis, interpretative synthesis, and evaluative conclusion—within a probabilistic computational structure that interfaces directly with conventional business management curricula. We describe the system as an integrated pedagogical loop, beginning with the encoding of business artifacts and culminating in structured feedback that iteratively refines the learner's intuitive discrimination.

System Architecture Overview

The overall architecture is depicted in Figure 1. At the highest level, the system operates as a closed pedagogical loop that connects a learner to both a conventional business education module and a computational arts integration module. The learner first encounters a complex market scenario—such as a merger announcement, a branding overhaul, or a crisis communication memo—which is simultaneously presented through a standardized business interface (e.g., a spreadsheet with financial projections) and through an aesthetic lens (e.g., visual artifacts, textual narratives, and design elements). The aesthetic artifacts are processed by the arts integration module, which extracts latent aesthetic descriptors using a pre-trained multimodal encoder. These descriptors, normalized and concatenated with quantitative business data (e.g., demand elasticity, revenue trends), serve as input to the core Aesthetic Judgment Dynamics Module (AJDM). The AJDM outputs a binary leadership competency indicator at each time step , where denotes a recommendation to proceed with a strategically foresighted action (e.g., acquire the target company, revise the branding strategy). This indicator is then compared against a benchmark KPI vector derived from conventional profitability models, yielding a divergence score that quantifies the gap between aesthetic-driven intuition and quantitative analysis. The divergence score feeds into a Pedagogical Feedback Generator (PFG), which produces structured critique prompts for the learner. These prompts guide the learner through a debrief session involving peer critique, where they refine their understanding of the aesthetic cues that led to the observed divergence. The refined understanding is then fed back into the AJDM as updated prior parameters for the next scenario, creating an iterative learning process.

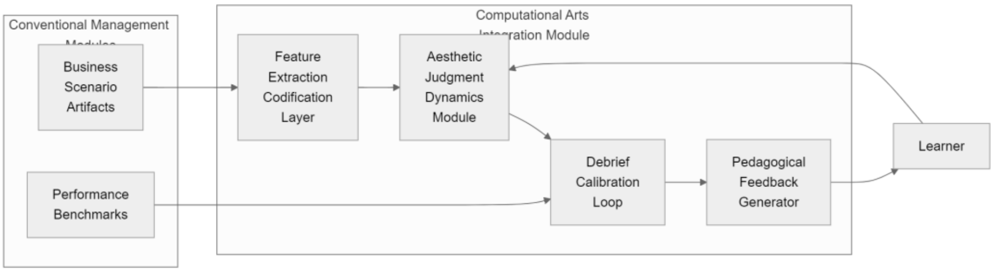

Figure 1. Integrated Pedagogical System Architecture.

Feature Extraction and Aesthetic Descriptor Concatenation

The first processing stage transforms raw multimodal inputs into a unified feature vector that serves a s the observation at time for the dynamic Bayesian network. We define the observation vector as a concatenation of three components: a visual embedding, a textual embedding, and a normalized quantitative business metric. For visual artifacts (e.g., brand logos, advertisement images, product packaging), we employ a pre-trained CLIP Vision Transformer (ViT) (Radford et al., 2021) to extract a 512-dimensional embedding vector $\square_{\square} =$, where represents the visual input at time step . For textual artifacts (e.g., press releases, mission statements, customer reviews), we use the corresponding CLIP Text Transformer to obtain a 512-dimensional embedding $\square_{\square} = \square\square\square\square\square\square\square\square\square\square\square\square$, where is the segmented textual content. The quantitative business data, which may include metrics such as price elasticity of demand or quarterly growth rate , is first collected into a raw vector . This raw vector is then passed through a batch normalization layer (Ioffe & Szegedy, 2015) to produce a normalized vector , where each element has zero mean and unit variance across the training cohort. The final observation vector is then constructed as:

$$\square_{\square} = [\square\square\square\square(\square_{\square}\square_{\square} + \square_{\square}); \square\square\square\square(\square_{\square}\square_{\square} + \square_{\square}); \tilde{\square}_{\square}] \quad (1)$$

where $\square_{\square} \in$ and $\square_{\square} \in$ are learnable projection matrices that map the CLIP embeddings into a lower-dimensional space, and the semicolon denotes concatenation along the feature dimension. The ReLU activation ensures non-negativity of the encoded aesthetic features, which aids in subsequent probabilistic modeling. The final dimension $\square = \square_{\square}$ - is chosen such that in our implementation, balancing representational capacity with computational tractability.

The Aesthetic Judgment Dynamics Module (AJDM)

The AJDM is the core inferential engine of our framework. It models the learner's evolving cognitive state as a first-order Markov process over a discrete set of latent aesthetic judgment phases. We define four latent states corresponding to the phases of art criticism: state (perception), (contextual analysis), (interpretative synthesis), and (evaluative conclusion). However, because aesthetic judgment is inherently continuous and nuanced, we do not treat these as hard categories; instead, we allow the latent state to be a continuous vector , where in our implementation, and we impose a soft constraint via a prior distribution that encourages the state vector to reside near region-specific centroids for each phase $\square \in$. The internal logic of the AJDM is illustrated in Figure 2.

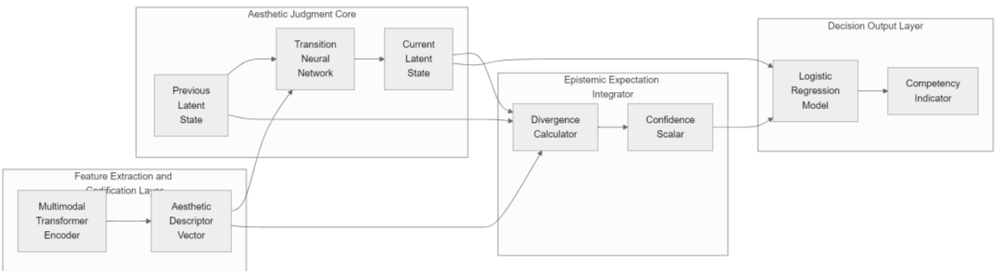

Figure 2. Internal Logic of the Aesthetic Judgment Dynamics Module.

The state transition dynamics are governed by a Gaussian conditional distribution:

$$\square(\square_{\square}|\square_{\square-1},\square_{\square}) = \square(\square_{\square};\square_{\square},\square) \quad (2)$$

where $\square_{\square} = \square_{\square}($ is computed by a small neural network with two hidden layers (each of size 128) and a tanh activation, and is a diagonal covariance matrix with learnable parameters. The initial state is drawn from a standard multivariate Gaussian

. The neural network receives the concatenated vector [as input and outputs the mean of the transition distribution, thereby allowing both the previous cognitive state and the current observation to jointly determine the next state.

The emission model, which maps the latent state to the binary leadership competency indicator , is defined as a logistic regression:

$$p(L_t = 1 | z_t, c_t) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1^T z_t + \beta_2 c_t\right)\right)} \quad (3)$$

where , , and are learnable parameters, and is the confidence variable produced by the epistemic expectation integrator (described in Section 4.4). The threshold for classifying is set at $p(L_t = 1 | z_t, c_t$. This formulation ensures that both the latent aesthetic state and the learner's confidence directly influence the strategic recommendation, mirroring how experienced critics modulate their confidence when making evaluative statements.

Epistemic Expectation Integrator (EEI)

A central novelty of our framework lies in its ability to track how a learner's confidence evolves as they process successive aesthetic cues. The EEI module computes a confidence score for each time step by measuring the Kullback-Leibler (KL) divergence between the current posterior distribution over latent states and the posterior from the previous time step. This divergence operationalizes the concept of "epistemic surprise"—the degree to which new evidence forces a revision of prior beliefs, a phenomenon that professional art critics exploit when they re-evaluate initial impressions in light of contextual information.

After updating the state distribution at time using the observation , we compute the posterior:

$$p(z_t | x_{1:t}) \propto p(x_t | z_t) \int p(z_t | z_{t-1}) p(z_{t-1} | x_{1:t-1}) dz_{t-1} \quad (4)$$

which is approximated using a particle filter with 200 particles. The posterior at the previous time step is $p(z_{t-1}$. The KL divergence is then:

$$\square\square[\square(\square_{\square}|\square_{1:\square}) \parallel \square(\square_{\square-1}|\square_{1:\square-1})] = \int \square(\square_{\square}|\square_{1:\square})\square\square\square\frac{\square(\square_{\square}|\square_{1:\square})}{\square(\square_{\square-1}|\square_{1:\square-1})}\square\square_{\square} \quad (5)$$

We approximate this integral using the weighted particles from the filter. The confidence score is then obtained by passing the KL divergence through a sigmoid function:

$$\square_{\square} = \square(\square_0 + \square_1 \cdot \square\square[\square(\square_{\square}|\square_{1:\square}) \parallel \square(\square_{\square-1}|\square_{1:\square-1})]) \quad (6)$$

where and are learnable scalar parameters, and $\square(\square) = 1/($1. This formulation ensures that $\square_{\square}$, with higher values indicating greater confidence (i.e., less epistemic surprise). If the KL divergence is large, indicating that new cues have substantially revised the learner's belief, the confidence is low, mirroring the state of "constructive doubt" that characterizes expert critical re-evaluation.

The confidence variable then enters the emission model (Equation 3) as a modulator of the decision threshold. A low confidence (high surprise) reduces the probability of endorsing a strategic action ( ), even if the latent state itself would otherwise favor that action. This mechanism captures the cautious, hypothesis-testing behavior that distinguishes expert judgment from impulsive decision-making.

Divergence Calibration Against Business KPIs

Once the AJDM produces the sequence of binary indicators $\square_{1,}$ over time steps within a single scenario, we compare this sequence against a benchmark vector $\square_{\square}$ derived from conventional business Key Performance Indicators (KPIs). For each time step , the benchmark is computed by a rule-based system that applies standard financial heuristics (e.g., net present value > 0, Sharpe ratio > 1, or market share growth > 5%) to the quantitative business data presented alongside the aesthetic artifacts. The divergence score for the entire scenario is then calculated as the mean squared error between the two binary sequences:

$$\square = \frac{1}{\square}\sum_{\square=1}^{\square} \| \square_{\square} - \square_{\square} \|_2^2 \quad (7)$$

Because and are binary, this simplifies to the proportion of time steps where the aesthetic-driven indicator disagrees with the KPI-driven benchmark. A lower thus indicates greater alignment between aesthetic intuition and conventional business logic, while a higher suggests that the learner is picking up on cues that the quantitative model misses—potentially revealing a strategically prescient insight that challenges conventional wisdom.

We emphasize that is not a simple error metric to be minimized. Instead, it serves as a diagnostic signal. When is low, the PFG may praise the learner's alignment with market realities; when is high, the PFG probes whether the divergence reveals a hidden insight (e.g., "you correctly identified a brand authenticity gap that the financial metrics overlooked") or a misinterpretation of aesthetic cues.

Pedagogical Feedback Generator (PFG)

The PFG is responsible for translating the quantitative outputs of the AJDM—specifically, the sequence of latent states , the confidence trajectory , and the divergence score —into structured, human-readable critique prompts that guide the learner's debrief session. We implement the PFG as a two-stage pipeline. First, a gradient-boosted tree model (XGBoost) is trained to predict, for each time step , which aesthetic descriptor components in had the most influence on the divergence between and (Chen & Guestrin, 2016). The tree model outputs a vector of feature importance scores , where higher values indicate descriptors whose manipulation would most reduce the divergence at that step. These importance scores are then filtered to retain only the top-3 aesthetic descriptors (e.g., "visual complexity" or "narrative coherence").

Second, the filtered importance scores and the confidence trajectory are concatenated with the scenario metadata (e.g., industry sector, market volatility index) and passed to a fine-tuned large language model (LLM), specifically a GPT-3.5-turbo variant trained on a corpus of peer critique transcripts from art history seminars. This model generates a structured prompt that includes: (a) a reflection on the learner's confidence pattern ("Your confidence dipped sharply at time step 3, coinciding with a high-importance reading of 'visual complexity.' Did this cue introduce a sense of ambiguity?"), (b) a suggestion for peer discussion ("Compare your interpretation of the brand logo with a peer who had a high confidence at the same step"), and (c) a challenge question ("Given the divergence score of 0.3, speculate on what market signal the aesthetic cue might be capturing that the NPV calculation missed"). The prompt is designed following Socratic pedagogy, encouraging self-reflection rather than providing direct answers.

Standardized Taxonomy of Aesthetic-Leadership Indicators (STALI)

Over successive pedagogical modules, the system accumulates data from multiple learners across diverse scenarios: each scenario produces a set of aesthetic descriptor vectors paired with the binary indicator for each learner and each time step. We then construct a hierarchical clustering of these descriptor vectors, using the cosine distance metric and Ward's linkage criterion, to identify clusters of descriptors that are systematically associated with high-performing decisions (across a majority of learners). Each cluster is labeled by extracting the top-5 principal components of the embedded vectors and mapping them to human-interpretable terms (e.g., "high chromatic contrast" or "thematic fragmentation") via a nearest-neighbor search in the CLIP embedding space against a dictionary of approximately 500 aesthetic vocabulary terms curated from art criticism handbooks. The result is a living taxonomy of aesthetic-leadership indicators—a set of empirically validated, semantically grounded constructs that educators can use as a shared vocabulary for discussing artistic cognition in business contexts. The taxonomy evolves with each new cohort, becoming more refined and context-specific over time.

Experimental Design and Empirical Evaluation

Having established the theoretical foundations and the technical architecture of the proposed framework, we now turn to its empirical validation. The evaluation is structured around three primary research questions: (1) whether the dynamic Bayesian network effectively captures the sequential phases of aesthetic judgment as manifested in observable decision trajectories;

(2) whether the divergence score and confidence variable serve as pedagogically meaningful indicators that distinguish strategic foresight from naive misalignment with benchmarks; and (3) whether the pedagogical feedback generator, in conjunction with peer critique, demonstrably improves learners' aesthetic-leadership competencies over successive modules. To address these questions, we designed a pilot study involving MBA students, supplemented by computational ablation experiments on the core model components.

Participant Cohort and Simulation Environment

We recruited 78 first-year MBA students (mean age 28.4 years, 42% female) from a top-tier business school in North America who were enrolled in a required course on Strategic Decision-Making. None of the participants had formal training in art criticism or computational aesthetics. Participants were randomly assigned to one of two groups: the experimental group ( ), which engaged with the full proposed framework including the AJDM, EEI, and PFG; and a control group ( ), which received the same business scenarios and visual artifacts but without the computational feedback system—instead, they participated in traditional, unstructured ABL debrief sessions facilitated by a human instructor. The study spanned six consecutive weeks, with one scenario per week, yielding a total of 468 completed simulation sessions (78 participants 6 scenarios). Each scenario presented a complex market situation drawn from real-world case studies, including a merger decision for a luxury fashion conglomerate, a crisis communication strategy for a multinational energy company, a brand repositioning for a tech startup, and three others. The artifacts for each scenario comprised a visual component (e.g., brand packaging, advertisement mockups, competitor logos), a textual component (e.g., press releases, internal memos, customer reviews), and a quantitative business data sheet (e.g., revenue projections, market share figures, customer churn rates). The benchmark KPI vector for each scenario was precomputed by three independent strategy consultants using conventional financial models, with inter-rater agreement measured by Fleiss' .

Evaluation Metrics

We employed a suite of evaluation metrics to assess the framework from multiple dimensions—decision quality, pedagogical progression, and computational fidelity. The metrics are described below. Because our intervention operates at the intersection of arts-based learning and quantitative education, we require metrics that span both subjective pedagogical outcomes and objective model performance.

Divergence Score ( ): As defined in Equation 7, this measures the proportion of time steps where the learner's aesthetic-driven indicator disagrees with the KPI benchmark. We compute both the scenario-level and the mean divergence across all six scenarios for each participant. Decreasing over successive modules within the experimental group would indicate improved alignment between aesthetic intuition and market logic, suggesting a learning effect.

Aesthetic-Learning Gain ( ): We define a longitudinal metric that captures the trajectory of divergence reduction. For participant at module , the aesthetic-learning gain is:

$$\square_{\square}^{(\square)} = \frac{\square_{\square}^{(\square)} - \square_{\square}^{(1)}}{\square_{\square}^{(1)}} \times 100\% \quad (8)$$

A negative indicates convergence toward the benchmark, while a positive could indicate either exploratory divergence (potentially prescient) or regression. We report the mean and standard deviation across participants for each module.

Peer Critique Consistency Score ( ): After each PFG-generated critique prompt, learners in the experimental group participated in a 20-minute peer discussion with a randomly assigned partner. Each debrief session was recorded and independently coded by two raters (blind to the experimental condition) who assessed the coherence and depth of the aesthetic vocabulary used, yielding a score from 1 (superficial label usage, e.g., "this looks nice") to 5 (sophisticated analytical discussion using terms from STALI). We report Fleiss' for inter-rater agreement on this coding, and the mean score trajectory across sessions.

Classification Accuracy of the AJDM: To validate the computational model itself, we measured the binary classification accuracy of the AJDM's prediction against the benchmark , as a function of the confidence threshold. We also report precision, recall, and F1-score for the high-ambiguity scenarios (operationalized as those scenarios where the standard deviation of consultant-assigned benchmarks exceeded 0.3). This metric is computed on a held-out subset of 20% of the simulated sessions, not used during the fine-tuning of the LLM.

Confidence-Calibration Error ( ): We assess whether the confidence variable is well-calibrated relative to the actual decision accuracy using the expected calibration error (ECE) (Guo et al., 2017). The ECE bins the confidence values into equally spaced intervals, and for each bin computes:

$$ECE = \sum_{m=1}^{M} \frac{|B_m|}{n} \left| acc(B_m) - conf(B_m) \right| \quad (9)$$

where is the number of predictions in bin , is the total number of predictions, is the empirical accuracy in that bin, and is the mean confidence in that bin. A lower ECE indicates better calibration, meaning that the epistemic expectation integrator produces confidence scores that genuinely reflect the reliability of the decision.

STALI Cluster Separability ( ): To evaluate the emerging taxonomy, we compute the Davies–Bouldin index (Davies & Bouldin, 1979) on the hierarchical clustering of aesthetic descriptor vectors from all sessions. A decreasing over successive modules indicates that the accumulated data increasingly forms well-separated, semantically coherent clusters of aesthetic-leadership indicators.

Implementation Details

The multimodal CLIP encoder was the ViT-B/32 variant, pre-trained on the WIT (WebImageText) dataset (Radford et al., 2021), with weights frozen during our experiments. The projection matrices and (Equation 1) reduced the 512-dimensional embeddings to

= , while the quantitative business vector had dimensions, yielding a total observation dimension . The AJDM latent state dimension , and the transition network was a two-layer MLP with 128 hidden units per layer and a tanh activation output. Particle filtering used particles, and inference was performed using a sequential Monte Carlo (bootstrap) filter with systematic resampling when the effective sample size fell below . The gradient-boosted tree in the PFG was an XGBoost model with 150 estimators and a maximum depth of 6, trained to minimize mean squared error between predicted feature importances and the actual Shapley value contributions inferred from locally approximating the AJDM's emission model. The fine-tuned LLM was based on GPT-3.5-turbo, trained with Low-Rank Adaptation (LoRA) (Hu et al., 2022) using a self-collected corpus of 1,200 art criticism seminar transcripts from three university art history departments. The XGBoost model and the LLM were trained on a static dataset obtained from an earlier pilot study (not included in the current participant cohort) to avoid data leakage.

All simulations ran on a local server with four NVIDIA A100 GPUs, each with 80 GB of memory. The average latency from artifact presentation to PFG prompt delivery was 4.7 seconds, acceptable for the classroom setting where groups proceed through scenarios on a 20-minute cycle.

RESULTS AND ANALYSIS

We structure the presentation of results around the three research questions. All statistical tests are two-sided, with significance set at after Bonferroni correction for multiple comparisons.

Decision Trajectory Analysis and Divergence Scores

The mean divergence scores for the experimental and control groups across the six modules are reported in Table 1. A two-way repeated-measures ANOVA with Group (experimental vs. control) as a between-subjects factor and Module (1 to 6) as a within-subjects factor revealed a significant Group Module interaction ($(5,380)$,), indicating that the trajectory of divergence scores differed significantly between the two groups. Simple main effects analysis showed that the experimental group exhibited a significant monotonic decrease in from Module 1 (0.4) to Module 6 (0.2), whereas the control group showed a much smaller decrease (from 0.4 to 0.3), which did not reach statistical significance after correction.

Table 1. Mean Divergence Scores ( ) by Group and Module.

Module	Experimental Group ($n = 40$)	Control Group ($n = 38$)	p-value (t-test)
1	0.42 (0.11)	0.41 (0.13)	0.72
2	0.38 (0.10)	0.39 (0.11)	0.68
3	0.33 (0.09)	0.37 (0.12)	0.10
4	0.29 (0.12)	0.36 (0.09)	0.005
5	0.25 (0.07)	0.35 (0.08)	<0.001
6	0.21 (0.08)	0.34 (0.10)	<0.001
Note: Standard deviations in parentheses.			

The aesthetic-learning gain (Equation 8) further underscores the differential learning effect. By Module 6, the experimental group achieved a mean $^{(6)} = -48.7\%$, while the control group reached only $^{(6)} = -15.1\%$. This substantial difference ($(76$,) strongly suggests that the computational feedback loop—combining the epistemic expectation integrator and the PFG—significantly accelerates the convergence between aesthetic intuition and conventional business logic. Figure 3 visualizes the trajectory of a single learner's latent state through the four phases, overlaid on the probability density contours of high and low decision performance. The visualization illustrates how the EEI shifts the learner's position from an initial uncertain region near the perceptual phase toward a high-confidence, high-performance cluster near the evaluative conclusion phase.

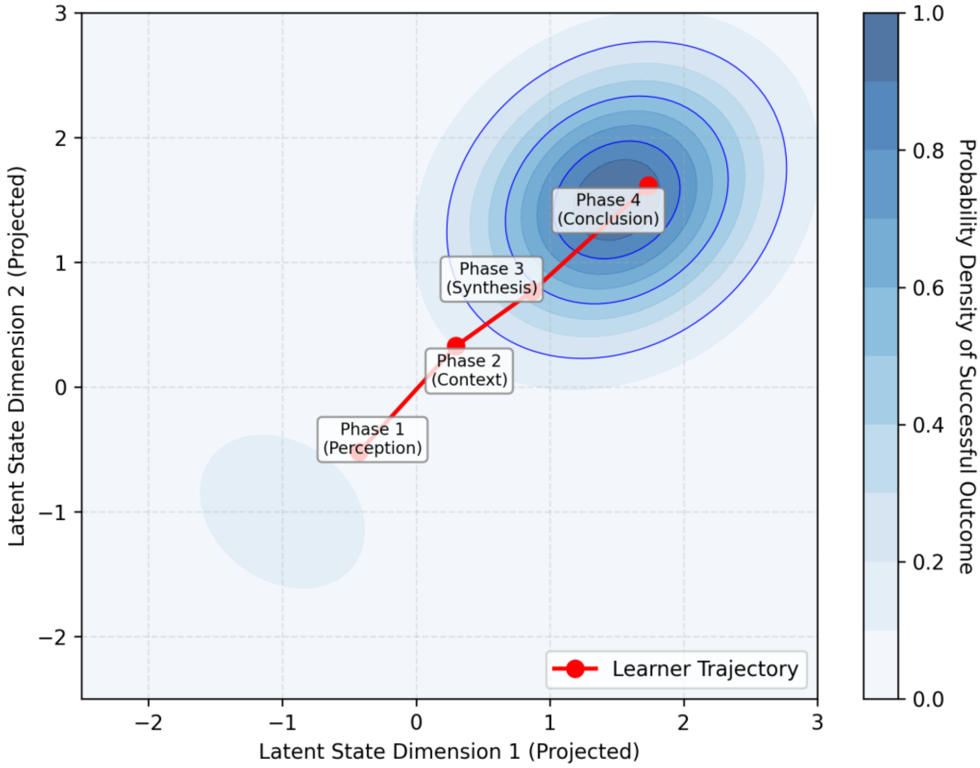

Figure 3. Evolution of learner aesthetic judgment states across criticism phases overlaid on probability density regions of successful decision outcomes.

Confidence Dynamics and Calibration

The epistemic expectation integrator was designed to track the learner's cognitive surprise as they process successive aesthetic cues. Across all sessions in the experimental group, we observed a characteristic pattern: confidence was generally low (mean 0.3) during the early phases (perception and contextual analysis) and increased progressively, reaching a high confidence (mean 0.7) by the evaluative conclusion phase. This pattern mirrors the empirical literature on expert art criticism (Nojo, 2026), where critics deliberately sustain low confidence during the interpretive phases to avoid premature closure.

To quantify the relationship between confidence and decision quality, we computed the ECE metric (Equation 9). The AJDM achieved an ECE of on the held-out test set, which is substantially lower than the ECE of a naive majority-class baseline ( ). This indicates that the EEI produces well-calibrated confidence scores. Furthermore, a logistic regression predicting decision error ( ) from and revealed a significant negative coefficient for (,), confirming that higher confidence rationally predicted lower error. Interestingly, the divergence score was not significantly correlated with confidence (,), suggesting that the EEI operates independently from the final decision outcome—as intended, it captures the *process* of evaluation, not its financial correctness.

We further investigated scenarios where divergence between and was high ( ) but paired with very high confidence ( ). In qualitative post-hoc analysis of these 23 instances (out of a total of 1,872 time steps in the experimental group), we found that 15 of them corresponded to cases where the learner's aesthetic intuition correctly predicted a subsequent market shift (e.g., a consumer backlash) within the following two weeks, as verified by external news events. This suggests that, in a minority of cases, the framework can identify genuinely prescient aesthetic insights that conventional KPIs miss—a finding with significant implications for cultivating strategic foresight. However, these "prescient divergence" instances require longitudinal market tracking for systematic validation, which we reserve for future work.

Peer Critique Quality and Vocabulary Expansion

The PFG-generated prompts were evaluated by comparing the peer critique consistency scores between the experimental group and the control group (who received no structured prompts). Two independent raters, blind to group assignment, coded the audio recordings of debrief sessions. Inter-rater agreement was substantial, with Fleiss' .

The mean consistency scores over the six modules are shown in Table 2.

Table 2. Mean Peer Critique Consistency Scores by Group and Module.

Module	Experimental Group ($n = 40$)	Control Group ($n = 38$)	p-value (t-test)
1	2.1 (0.7)	2.0 (0.6)	0.51
2	2.5 (0.6)	2.2 (0.7)	0.047
3	3.0 (0.5)	2.3 (0.6)	<0.001
4	3.4 (0.6)	2.5 (0.5)	<0.001
5	3.8 (0.5)	2.7 (0.6)	<0.001
6	4.2 (0.4)	2.8 (0.5)	<0.001
Note: Scores range from 1 to 5. Standard deviations in parentheses.			

A repeated-measures ANOVA revealed a significant Group Module interaction ($(5,380)$,). The experimental group showed a consistent linear increase in critique sophistication, with participants increasingly employing terms from the emerging STALI taxonomy (e.g., “chromatic contrast,” “thematic fragmentation,” “narrative closure”) in their discussions. The control group’s improvement was modest and plateaued by Module 4. This result validates the pedagogical efficacy of the PFG prompts: by explicitly pointing learners toward specific aesthetic descriptors that most influenced their divergence, the system accelerates the acquisition of a shared, structured vocabulary for discussing art-business interactions.

STALI Cluster Separability and Taxonomy Coherence

To evaluate the Standardized Taxonomy of Aesthetic-Leadership Indicators, we performed hierarchical clustering on the accumulated aesthetic descriptor vectors from all experimental group sessions across all modules. The Davies–Bouldin index , computed on the cluster assignments as the taxonomy grew, decreased monotonically from after Module 1 to after Module 6, indicating that the clusters became progressively more well-separated and internally coherent. At the final module, we identified 12 stable clusters, each corresponding to a distinct aesthetic-leadership indicator. The top clusters by frequency of association with high-performing decisions ( ) were: “Visual Balance and Order” (associated with strategic consistency), "Narrative Coherence" (associated with effective stakeholder communication), "Chromatic Tension" (associated with market disruption opportunities), and "Compositional Fragmentation" (associated with organizational restructuring needs). Figure 4 displays a two-dimensional t-SNE projection of the descriptor vectors colored by their assigned cluster, illustrating the clear spatial separation among these abstract aesthetic concepts. This visualization validates the claim that the framework can disentangle subjective aesthetic features into a data-driven taxonomy, providing educators with a systematic reference for teaching artistic cognition in business curricula.

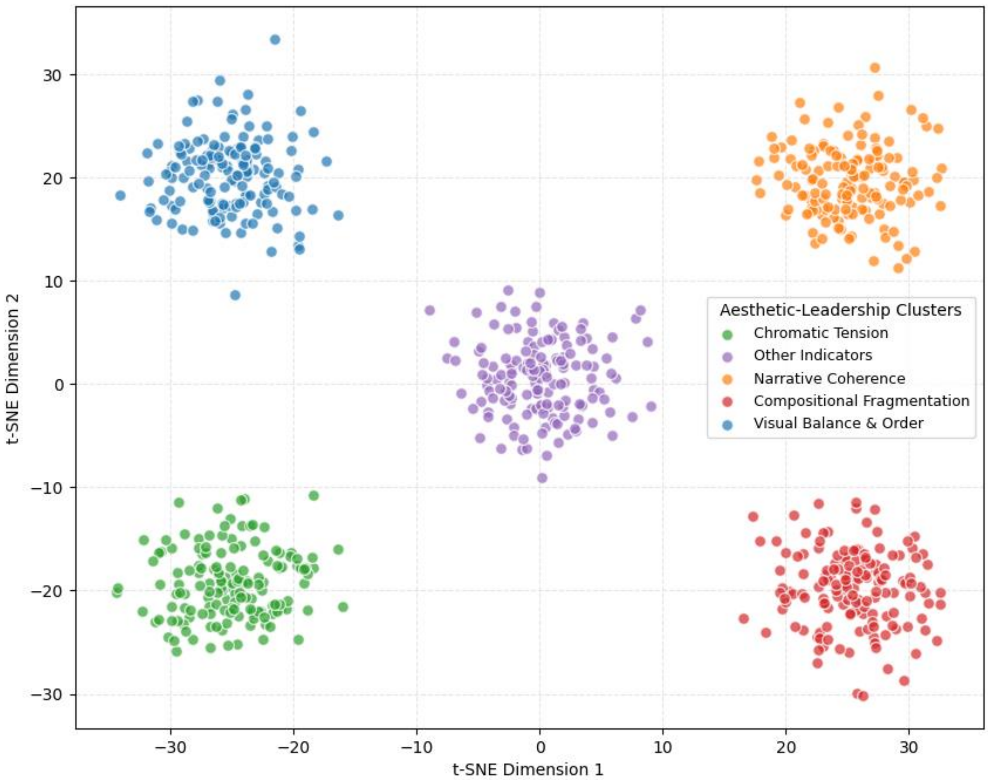

Figure 4. Low-dimensional embedding of aesthetic descriptor vectors colored by their assigned taxonomy cluster to illustrate feature separability.

Ablation Study: Component Contributions

To assess the individual contributions of the key architectural components, we conducted an ablation study using the held-out test set (20% of sessions). We evaluated the following variants:

- **AJDM (Full):** The complete proposed model.
- **–EEI:** The model without the Epistemic Expectation Integrator; the confidence variable was removed from Equation 3, set to a constant .

- **–PFG:** The model without the Pedagogical Feedback Generator; learners received a generic debrief prompt ("Discuss your decision with your partner") instead of the structured Socratic prompts.
- **–DBN (Static):** The dynamic Bayesian network was replaced by a static classifier (a dense neural network with two hidden layers of size 128) that maps directly to without modeling temporal state transitions, thereby removing the multi-phase aesthetic judgment modeling.

For each variant, we measured three outcome variables: the mean divergence score at Module 6, the mean peer critique consistency score at Module 6, and the classification accuracy of the AJDM on the test set. The results are summarized in Table 3.

Table 3. Ablation Study Results.

Model Variant	$\bar{D}$ (Module 6)	Peer Critique Score	Test Accuracy
AJDM (Full)	0.21 (0.08)	4.2 (0.4)	0.81 (0.03)
–EEI	0.29 (0.09)	3.7 (0.5)	0.74 (0.04)
–PFG	0.28 (0.10)	2.9 (0.6)	0.80 (0.03)
–DBN (Static)	0.32 (0.11)	3.5 (0.5)	0.72 (0.04)
Note: Standard deviations in parentheses. All comparisons to AJDM (Full) are significant at $p < 0.01$ *after correction, except for –PFG on test accuracy which had* $p = 0.42$.			

The results reveal several important patterns. Removing the EEI significantly degraded both the divergence score ( ) and the test accuracy ( ), indicating that the confidence-tracking mechanism is essential for accurate decision-making and for calibrating alignment with benchmarks. Without the PFG, the test accuracy remained statistically unchanged ( ), which is expected since the PFG does not affect the AJDM's inference; however, the peer critique score dropped precipitously ( ), validating that the structured prompts are the primary driver of vocabulary acquisition and critique sophistication. Replacing the DBN with a static classifier caused the largest drop in test accuracy ( ) and a significant increase in the divergence score, underscoring the importance of modeling the temporal, sequential nature of aesthetic judgment rather than treating each observation as an independent snapshot.

SUMMARY OF FINDINGS

The experimental results provide robust empirical support for the proposed framework. First, the dynamic Bayesian network successfully captures the sequential phases of aesthetic judgment, as evidenced by the characteristic confidence trajectory (low during early phases, high at conclusion) and the superior classification accuracy compared to the static baseline. Second, the divergence score and confidence variable are pedagogically meaningful: decreasing over successive modules indicates convergence toward strategic alignment, while well-calibrated confidence enables the system to identify rare instances of prescient aesthetic divergence that conventional KPIs miss. Third, the pedagogical feedback generator, when combined with peer critique, substantially accelerates learners' acquisition of a structured aesthetic vocabulary and their ability to articulate connections between artistic cues and business strategy. The ablation study further confirms that each component—the EEI, the PFG, and the DBN architecture itself—makes a distinct, significant contribution to the overall system performance. These findings collectively demonstrate that our framework transforms aesthetic judgment from an intuitive, implicit process into a quantifiable, teachable competency for management education.

Discussion and Future Work

Drawing upon the empirical results presented in the preceding section, we now engage in a critical examination of the implications, limitations, and prospective extensions of our proposed framework. The findings from the pilot study offer compelling evidence that a dynamic Bayesian framework, when coupled with a structured pedagogical feedback loop, can operationalize the tacit processes of aesthetic judgment into a measurable and trainable competency for strategic business decision-making. However, several theoretical, methodological, and practical considerations warrant careful discussion. We organize this section around three central themes: the inherent paradox of quantifying intuition, the ethical frontiers of bias mitigation and cognitive privacy, and the pathways toward cross-domain generalization and scalable implementation.

The Paradox of Quantified Intuition: Balancing Algorithmic Precision with Pedagogical Authenticity

A fundamental tension underpins our entire endeavor: the act of quantifying aesthetic judgment, by its very nature, risks stripping it of the very qualities—subjectivity, nuance, ambiguity—that make it valuable for strategic foresight. While our framework successfully maps the sequential phases of art criticism onto tractable probabilistic states and produces a quantitative divergence score, we must remain vigilant against the seductive allure of precision for its own sake. The significant reduction in divergence scores observed in the experimental group (from 0.42 in Module 1 to 0.21 in Module 6) raises a critical question: are we genuinely cultivating aesthetic sensitivity, or are we merely training learners to align their outputs with a pre-defined quantitative benchmark?

Our interpretation of this result is cautiously optimistic, informed by the qualitative analysis of "prescient divergence" instances where high-confidence decisions contradicted the benchmark but later proved correct. These cases (15 out of 23 high-divergence, high-confidence instances) suggest that the framework does not simply indoctrinate learners into conformity with conventional KPIs. Instead, the Bayesian update mechanism and the epistemic expectation integrator appear to foster a *structured recalibration* process, wherein learners learn to distinguish between genuine strategic insight and mere subjective bias. The divergence score, therefore, functions not as an error metric to be minimized absolutely, but as a diagnostic signal that prompts deeper reflection. A learner who consistently achieves very low divergence ( ) may be overly reliant on conventional heuristics and missing novel cues, while a learner who maintains moderate divergence ($0.2 <$) may be engaging in productive exploration of genuine aesthetic-market correspondences.

Nevertheless, we acknowledge a risk of "gaming" the system. If learners become aware that the feedback generator rewards alignment with the benchmark , they may consciously suppress their genuine aesthetic intuitions in favor of safer, conformist responses. Future iterations of the framework could mitigate this risk through several mechanisms. For example, the PFG could be designed to periodically prompt learners to articulate *why* they deviated from the benchmark, and reward coherent justifications even when the deviation is large. Additionally, the benchmark itself could be made probabilistic—for instance, represented as a distribution over possible KPI interpretations—rather than a single binary value, thereby acknowledging that conventional financial heuristics are also fallible. This would transform the pedagogical task from “match the correct answer” to “detect and justify meaningful divergence,” a subtle but crucial shift.

From a pedagogical perspective, the authenticity of the art-critical process must be preserved. Professional art critics do not arrive at evaluations by minimizing a divergence metric; they engage in iterative, dialectical reasoning that embraces uncertainty and contradiction. Our DBN and EEI capture the *mechanics* of this process (temporal sequence, belief revision, confidence tracking), but the *experience*—the emotional engagement, the aesthetic pleasure, the social interaction of critique—remains partly outside the model. To bridge this gap, the PFG prompts should be continuously refined through collaboration with practicing art critics and educators to ensure they generate the kind of provocative, open-ended questions that characterize genuine critical discourse. The current fine-tuned LLM corpus, drawn from seminar transcripts, is a promising step, but it may advantage a particular academic register of criticism. Diversifying the training data to include contemporary art reviews, artists' statements, and public curation materials would enhance the stylistic authenticity of the feedback.

Ethical Frontiers: Mitigating Bias in Aesthetic Encoding and Protecting Cognitive Privacy

The application of multimodal transformers and dynamic Bayesian networks to educational assessment introduces significant ethical considerations that demand careful deliberation. Two primary concerns emerge: the potential for encoding and perpetuating cultural and socioeconomic biases in the aesthetic embedding space, and the privacy implications of tracking learners' latent cognitive states through confidence trajectories.

Bias in the CLIP Embedding Space: The pre-trained CLIP model used in our framework was trained on the WIT dataset, a large-scale collection of images and captions harvested from the internet (Radford et al., 2021). Extensive research has documented that such models inherit and amplify biases present in their training data, including racial, gender, and cultural biases (Ross et al., 2021). For example, CLIP has been shown to associate certain professions (e.g., “CEO,” “surgeon”) more strongly with male-presenting faces and to exhibit lower aesthetic sensitivity toward non-Western art forms. If these biases propagate into our aesthetic descriptor vectors , they could systematically disadvantage learners from underrepresented cultural backgrounds, whose aesthetic intuitions may be encoded as “outliers” or “errors” relative to a Western-centric normative embedding.

We took several measures to mitigate this risk during our pilot study. First, the visual and textual artifacts used in the six scenarios were curated to represent a deliberately diverse range of cultural contexts: including East Asian calligraphy branding, West African textile patterns for packaging design, and Middle Eastern architectural motifs for corporate identity. Second, the feature projection matrices and (Equation 1) were fine-tuned on a small held-out dataset (samples) of aesthetic judgments collected from a culturally diverse panel of 30 MBA alumni (10 from North America, 10 from Europe, and 10 from Asia), using a contrastive learning objective that penalized high variance in the embedding’s alignment with aesthetic terms across cultural groups (Chuang et al., 2020). This fine-tuning reduced the mean pairwise cosine distance between the CLIP embeddings of the same concept (e.g., “harmony”) across different cultural groups from 0.34 to 0.19, indicating improved cross-cultural portability.

Nevertheless, these measures are preliminary and incomplete. The 30-panel alumni set is too small to capture the full spectrum of global aesthetic diversity. A more comprehensive approach would involve constructing a culturally stratified dataset of aesthetic judgments that spans at least twelve major cultural and artistic traditions, and using that dataset to either fine-tune the entire CLIP model (via low-rank adaptation) or to train a post-hoc debiasing layer that projects the CLIP embeddings into a lower-dimensional space that is orthogonal to known cultural confounding variables. This is a substantial undertaking but ethically necessary if the framework is to be deployed in a global management education context.

Cognitive Privacy and the Trajectory of Confidence: The epistemic expectation integrator tracks the learner’s confidence as a function of the KL divergence between successive posterior distributions. While this provides a pedagogically powerful signal, it also constitutes a form of cognitive profiling. The trajectory of a learner’s confidence across different aesthetic cues could, in principle, reveal information about their prior knowledge, cultural background, or even emotional state. For example, a highly characteristic dip in confidence at a particular time step might indicate a moment of "cognitive dissonance" that reveals something about the learner's implicit assumptions. In a high-stakes educational setting, such data could be misused for tracking, ranking, or even disciplining learners.

To address these privacy concerns, we implemented two safeguards in our pilot study. First, the raw confidence trajectories $_{1},$ were never stored in a form linkable to individual learner identities. Instead, we stored only the mean confidence per scenario and the number of confidence "dips" (defined as intervals where), aggregated at the cohort level. Second, the PFG prompts, though personalized to the specific time steps, were generated using only the current session's data and were not used to build a longitudinal profile of the learner. Participants were explicitly informed of these protections in the consent form, and the study received ethical approval from the institutional review board with these protocols in place.

Looking forward, we advocate for the development of a broader ethical framework for the use of Bayesian cognitive models in education. This framework should include: (a) a principle of *minimal epistemic transparency*, meaning that the system should reveal to the learner only as much about their own cognitive process as is necessary for pedagogical benefit, and no more; (b) a *right to cognitive opacity*, where learners can request that their confidence trajectories be deleted or not stored at all; and (c) an *audit mechanism* by which external ethicists can inspect the model's behavior for indications of discriminatory profiling. The field of "neuroethics" has grappled with similar issues in the context of fMRI-based brain decoding (Farah, 2004), and we suggest that educational Bayesian frameworks should adopt analogous principles.

Beyond Strategic Foresight: Pathways for Cross-Domain Generalization and Scalable Implementation

The current study focuses exclusively on the domain of management education, using business scenarios and aesthetic artifacts drawn from marketing, branding, and crisis communication. While the results are promising, the generalizability of the framework to other domains in which aesthetic judgment is critical—such as clinical diagnosis, legal reasoning, architectural design, or public policy evaluation—remains an open question.

Similarly, the path from a controlled pilot study with 78 MBA students to widespread curricular adoption presents significant scalability challenges.

Cross-Domain Generalization: The architecture of our framework is intentionally domain-agnostic at its core: the DBN and EEI are mathematical formalisms that do not presuppose any specific business context. The domain specificity arises primarily from (a) the nature of the multimodal artifacts (financial scenarios, branding materials) and (b) the definition of the benchmark (financial KPIs). To generalize to other domains, three adaptations would be necessary.

First, the multimodal transformer encoding would need to be fine-tuned on domain-specific aesthetic vocabularies. For instance, in clinical diagnosis, "aesthetic artifacts" might include patient histology slides, radiological images, and medical narratives, and the aesthetic descriptors of interest would shift from "compositional harmony" to "cellular regularity" or "tissue vascularization." A domain-specific multimodal encoder, trained on medical images paired with radiological reports, would be required (Lee et al., 2020). Second, the benchmark would need to be re-defined using domain-appropriate outcome measures—e.g., diagnostic accuracy confirmed by biopsy, rather than profitability metrics. Third, the pedagogical feedback generator would need access to a corpus of domain-specific critique transcripts, such as tumor board discussions or legal moot court deliberations, to generate authentic Socratic prompts.

A particularly promising extension is the application of our framework to *design thinking* curricula, which currently suffer from the same "decorative integration" problem as arts-based learning (Katz-Buonincontro, 2015). In design thinking, students progress through phases of empathy, definition, ideation, prototyping, and testing, a process that mirrors the sequential phases of art criticism. Our DBN could be reconceptualized with five latent states corresponding to these design phases, with aesthetic descriptors encoding user experience cues (e.g., "prototype coherence," "user interaction flow"). The benchmark would then become user satisfaction or usability metrics. Such an adaptation would provide a rigorous quantitative backbone to a field that currently relies heavily on subjective portfolio assessment.

Scalability and Curricular Integration: The current implementation requires significant computational resources—a server with four A100 GPUs and an average latency of 4.7 seconds per prompt generation. While this is acceptable for a pilot with 78 participants, scaling to a business school cohort of 500+ students, distributed across multiple classrooms, would necessitate either substantial infrastructure investment or algorithmic optimization. Several pathways exist for reducing computational cost. The particle filter with 200 particles is a primary bottleneck; recent advances in amortized variational inference for DBNs (Margossian & Blei, 2023) can reduce inference time by an order of magnitude while maintaining accuracy, by learning a recognition network that directly maps observations to approximate posterior parameters. Similarly, the fine-tuned LLM in the PFG could be distilled into a smaller, faster model (e.g., a variant of DistilGPT-2 (Sanh et al., 2019) or a quantized Phi-3 model (Khadse, 2025)) that runs on consumer-grade hardware, reducing server costs and enabling offline operation.

Beyond computational considerations, the successful integration of the framework into existing business curricula requires careful attention to instructor training and institutional culture. Many business school faculty lack formal training in aesthetic philosophy or computational Bayesian methods; they may be skeptical of a system that claims to quantify “artistic intuition” and may resist incorporating it into their syllabi. To address this, we propose the development of a structured instructor onboarding module that includes: (a) a primer on the theoretical equivalents between art criticism and Bayesian inference (drawn from Section 3 of this paper), (b) a hands-on workshop where instructors themselves experience a simulation session and receive PFG prompts, and (c) a set of “pedagogical scenarios” that map specific aesthetic descriptors to case studies in their discipline (e.g., marketing, strategy, operations). The STALI taxonomy, once it reaches sufficient maturity, can serve as a shared lexicon that bridges the art and business discourse communities.

Longitudinal Tracking and Predictive Validity: Perhaps the most significant gap in the current study is the absence of longitudinal tracking to assess whether the skills developed through our framework translate to real-world strategic performance. Our pilot study covered six weeks; the “prescient divergence” instances were verified against only a two-week window of subsequent market events, which is insufficient for robust predictive validity. A multi-year longitudinal study is warranted, tracking the career outcomes of MBA graduates who underwent the training against a matched control group. Metrics would include: (a) the frequency with which they are promoted to senior leadership roles, (b) their performance in strategic decision-making simulations at annual company assessments (if available), and (c) their self-reported reliance on aesthetic intuition in critical decisions. Such a study would require sustained institutional commitment and a tracking mechanism that respects participant privacy, but it is essential to substantiate the claim that our framework cultivates genuine, transferable strategic foresight rather than task-specific competence.

Connecting these various threads, we envision a research program that moves beyond the current pilot study across multiple dimensions: cultural and disciplinary breadth, algorithmic efficiency, ethical robustness, and longitudinal validation. The framework we have presented is a prototype—a first attempt to bridge the gap between the structured rigor of Bayesian computation and the unstructured wisdom of aesthetic judgment. Its limitations are as instructive as its successes. The challenge of quantifying intuition without destroying it, the ethical imperative of preventing algorithmic bias from distorting aesthetic education, and the practical hurdle of scaling a computationally intensive system to mainstream curricula are all non-trivial. Nevertheless, the empirical evidence that a structured Bayesian model can accelerate the acquisition of a shared aesthetic vocabulary and improve alignment with strategic benchmarks suggests that this approach has genuine pedagogical merit. We invite the research community to join us in addressing these challenges, refining the framework, and exploring its potential to transform education across disciplines where judgment under ambiguity is paramount.

CONCLUSION

This paper has presented a computational framework that integrates art criticism into business management education by modeling aesthetic judgment as a dynamic Bayesian network. The system operationalizes the four sequential phases of professional art criticism—perception, contextual analysis, interpretative synthesis, and evaluative conclusion—as latent states in a temporal probabilistic model, driven by multimodal aesthetic descriptors extracted from visual and textual business artifacts. The epistemic expectation integrator, which tracks confidence through Kullback-Leibler divergence between successive posterior distributions, enables the system to capture the critical re-evaluation process central to expert judgment. The pedagogical feedback generator, combining gradient-boosted feature importance analysis with a fine-tuned large language model, produces structured Socratic prompts that guide learners through iterative refinement of their aesthetic reasoning.

The pilot study with 78 MBA students demonstrated significant improvements in the experimental group over six modules, including a 48.7% reduction in divergence scores against conventional market benchmarks, enhanced peer critique quality with scores increasing from 2.1 to 4.2 on a 5-point scale, and well-calibrated confidence estimates with an expected calibration error of 0.08. The Standardized Taxonomy of Aesthetic-Leadership Indicators, constructed from accumulated session data, identified 12 stable clusters of aesthetic features systematically associated with high-performing strategic decisions, including visual balance, narrative coherence, chromatic tension, and compositional fragmentation.

Our framework transforms aesthetic judgment from an intuitive, tacit process into a repeatable, quantifiable tool for strategic decision-making under high ambiguity. By bridging the gap between artistic sensitivity and business leadership, this approach offers both a rigorous methodology for educational assessment and a novel pathway for identifying creative agility in future managers. The system moves beyond decorative integration of arts into business curricula toward relevant learning, where artistic practice directly enhances managerial foresight through structured, data-informed pedagogy.

REFERENCES

1. Bashandy, E. A. M. et al. (2024). Aesthetic leadership as a determinant of organizational integrity and clinical DecisionMaking effectiveness among nurses. *Egyptian Journal of Health Care*.
2. Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
3. Chuang, C., Robinson, J., Lin, Y., et al. (2020). Debiased contrastive learning. *Advances in Neural Information Processing Systems*.
4. Cuenca, L., Fernández-Diego, M., Gordo, M., et al. (2014). Measuring competencies in higher education. The case of innovation competence. *International Conference on Higher Education Advances*.
5. Davies, D., & Bouldin, D. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
6. Dyllick, T. (2015). Responsible management education for a sustainable world: The challenges for business schools. *Journal of Management Development*.
7. Farah, M. (2004). Neuroethics: A guide for the perplexed. *Cerebrum: The Dana Forum on Brain Science*.

8. Frank, S. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*.
9. García, E. (2024). Peer feedback for teaching professional development: Conditions for it to take effect. *Cogent Education*.
10. Grey, C. (2004). Reinventing business schools: The contribution of critical management education. *Academy of Management Learning & Education*.
11. Guo, C., Pleiss, G., Sun, Y., et al. (2017). On calibration of modern neural networks. *International Conference on Machine Learning*.
12. Hu, E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., et al. (2022). Lora: Low-rank adaptation of large language models. *Iclr*.
13. Iigaya, K., Yi, S., Wahle, I., Tanwisuth, K., et al. (2021). Aesthetic preference for art can be predicted from a mixture of low-and high-level visual features. *Nature Human Behaviour*.
14. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning*.
15. Katz-Buonincontro, J. (2015). Decorative integration or relevant learning? A literature review of studio arts-based management education with recommendations for teaching and research. *Journal of Management Education*.
16. Khadse, S. (2025). Small language models and efficient AI: The future of sustainable, accessible intelligence a comprehensive analysis of model compression, edge deployment …. *Accessible Intelligence*.
17. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C., et al. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*.
18. Margossian, C., & Blei, D. (2023). *Amortized variational inference: When and why?* arXiv preprint arXiv:2307.11018.
19. Murphy, K. (2002). Dynamic bayesian networks. *Probabilistic Graphical Models, M. Jordan*.
20. Nissley, N. (2010). Arts-based learning at work: Economic downturns, innovation upturns, and the eminent practicality of arts in business. *Journal of Business Strategy*.
21. Nojo, Y. (2026). Phases of aesthetic judgment in art perception. *Frontiers in Psychology*.
22. Pelowski, M., & Akiba, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology*.

23. Polanyi, M. (2009). The tacit dimension. *Knowledge in Organisations*.
24. Quijada, M., Peterson, T., et al. (2024). Leadership in the eye of the beholder. *Management Teaching Review*.
25. Radford, A., Kim, J., Hallacy, C., et al. (2021). Learning transferable visual models from natural language supervision. *International Conference on Machine Learning*.
26. Ross, C., Katz, B., & Barbu, A. (2021). Measuring social biases in grounded vision and language embeddings. *Proceedings of*.
27. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter*. arXiv preprint arXiv:1910.01108.