International Journal on Semantic Web and Information Systems
Volume 22 • Issue 1 • January-December 2026

Evaluating Generative AI as a Triage Tool in Aligned Yet Divergent Investment Decision-Making

Yen-Hung Chen
National Taipei University of Nursing and Health Sciences, Taiwan

Chiao-Shan Chen
https://orcid.org/0000-0003-3852-6895
National Taipei University of Nursing and Health Sciences, Taiwan
Received: October 30th, 2025 | **Accepted:** January 8th, 2026

ABSTRACT

This study explores whether generative artificial intelligence (AI) can exhibit decision-making behavior aligned with that of human experts. A total of 200 startup projects were assessed across four key dimensions. Each project received parallel evaluations from investors and generative AI models. AI models aligned with human evaluators in overall score levels and moderately predicted human ratings yet differed substantially in their score distributions, and the contrasts between the top 20% and the bottom 80% segments across all three models further revealed a distinctly two-tier alignment structure. Two indicators showed the practical impact: human labor time decreased by 94–99.6%, and monetary cost per report dropped by 350–550 times. The results reveal general logic but missed expert-level nuance in bounded, gradual, and stratified alignment with expert evaluators. Bounded alignment reflects AI's structural limits, gradual alignment describes dimension-specific convergence with human judgment, and stratified alignment captures tiered patterns of human–AI co-evaluation.

KEYWORDS

Generative AI, Angel Investment Evaluation, Human–AI Alignment, Startup Incubator Decision-Making, Ranking Consistency and Scoring Divergence

INTRODUCTION

As generative artificial intelligence (GAI) becomes integrated into semantic web–oriented evaluation systems, a central challenge lies in determining whether ontology-based semantic reasoning can support automated scoring for large volumes of startup proposals (Abrusci et al., 2025; Elshani et al., 2025), thereby reducing the labor demands of preliminary review and compressing evaluation cycles from weeks to days (Albashrawi, 2025; Rossi, 2023). Prior studies highlight divergences between GAI and human evaluators in their scoring patterns, particularly when experts encounter unfamiliar domains driven by differing expectations of fairness, trust, and alignment in algorithmic decision-making (Chen, 2025; Dahlgren Lindström et al., 2025; Khamassi et al., 2024). Within this landscape, a critical unresolved question (and simultaneously the central aim of this study) is whether GAI-driven semantic triage can replicate, under domain-familiar review conditions in which evaluators specialize in specific proposal types instead of reviewing all categories, the conceptual structures

DOI: 10.4018/IJSWIS.399171

employed by human analysts and what systemic implications may arise if ontology-based automated scoring ultimately replaces manual evaluation (Osborne & Bailey, 2025).

Accordingly, the first problem demonstrates the lack of quantitative evidence on whether GAI-driven semantic triage can faithfully reproduce the evaluative schemas used by human experts (Zhang et al., 2025). Although AI systems sometimes align with human semantic structures, real-world deployments continue to display inconsistencies, indicating that alignment remains conditional rather than guaranteed. This gap underscores the need for systematic analyses that clarify when semantic alignment, ontological categorization, and metadata coherence between GAI and human evaluators converge and when they diverge. Addressing this issue may also reveal structural risks. Efficiency-oriented model calibration may compress ontological distinctions essential for detecting unconventional or culturally embedded proposals, such as expert differentiation between socially rooted microenterprise models and superficially similar standardized formats. Semantic compression reduces granularity in ontological categories, preventing the model from distinguishing culturally embedded or atypical cases, which increases misclassification risk by forcing diverse proposals into overly broad or homogenized conceptual structures. Moreover, even under identical prompting, variations in linguistic form and structural coherence across artificial intelligence (AI) outputs may undermine the semantic regularity required for mapping evaluations onto shared conceptual frameworks. Finally, it remains unclear whether clear boundaries between AI and human evaluators should be maintained, what risks emerge if such boundaries erode, and what systemic consequences follow if they are formally preserved (Osborne & Bailey, 2025; Zhang et al., 2025).

The first research question extends into a second, more structural inquiry: rather than documenting surface-level consistencies, we examine whether systematic convergence between GAI-driven semantic triage and human evaluators reflects a form of inbreeding, defined here as the recursive reinforcement of shared evaluative patterns across human and machine systems. We also investigate whether cognitive domestication emerges, which we operationalize as the bidirectional adjustment of human interpretive heuristics and AI-generated semantic frames during repeated interaction. Prior research shows that even non-anthropomorphic AI systems can trigger patterned domestication responses (Søraa & Fostervold, 2021); our study extends this logic to domain-familiar startup evaluation contexts. In these settings, human judgment patterns gradually converge toward centralized, unimodal distributions that resemble GAI outputs. We argue that such convergence indicates bidirectional domestication: GAI calibrated to approximate human conceptual structures begins shaping experts' expectations, while evaluators adapt their reasoning to the statistical regularities embedded in AI outputs, thereby stabilizing uncertainty within their decision processes. Simultaneously, this convergence suggests that institutional objectives may influence the semantic structure of AI-generated outputs, creating self-referential loops in human evaluative schemas (Gerstgrasser et al., 2024; Shumailov et al., 2024). These feedback dynamics risk reproducing institutional logics in ways that obscure political interests, incentive asymmetries, and conflicting priorities in real evaluation environments, ultimately encouraging unwarranted confidence in semantic automation.

Building on these insights, this study develops a quantitative, ontology-based analytical framework to assess how GAI and human experts co-produce semantic alignment. By integrating statistical analyses of 200 startup assessments with an ontology-driven reasoning architecture, the framework examines the conditions under which semantic triage approximates expert evaluative structures and the conditions under which their judgments diverge. This includes assessing how well AI reproduces overall scoring tendencies while identifying distributional differences, variation across risk-sensitive dimensions, and the emergence of stratified rather than continuous alignment patterns across the ranking distribution. In higher stakes regions (where evaluators rely on fine-grained distinctions), the framework tests whether alignment weakens, and in lower risk regions it evaluates whether broader structural cues enable more stable convergence. It also quantifies how shifts in linguistic form and structural coherence across AI outputs introduce volatility into semantic mapping. Further, the framework investigates how the boundary between AI and human evaluators is maintained or