

Computer-Aided Pitch Deck Pro: An Entrepreneurial Intelligence Framework for AI-Assisted Startup Pitch Generation, Investment Readiness Assessment, and Decision Support

Kristine T. Soberano

State University of Northern Negros, Philippines

Email: ksoberano@sunnn.edu.ph

ABSTRACT

The exponential growth of early-stage ventures coupled with the persistent scarcity of rigorous, scalable, and bias-aware investment-readiness assessment tools has produced a structural bottleneck in the global startup ecosystem. Although machine learning models now achieve over 90 % accuracy on retrospective success prediction and large language models can generate plausible business narratives, these capabilities remain siloed: predictive models offer no constructive feedback to founders, and generative models are routinely challenged by hallucination, weak grounding, and absence of structured evaluation rubrics. This paper introduces Computer-Aided Pitch Deck Pro, an integrated Entrepreneurial Intelligence Framework that operationalizes a seven-stage computational pipeline, Extract → Learn → Evaluate → Validate → Augment → Transform → Export, for end-to-end AI-assisted pitch deck generation, investor-readiness scoring, and decision support. Grounded in design-science research and a corpus-level synthesis of more than 120 peer-reviewed studies on venture AI, retrieval-augmented generation, explainable AI, and human-AI collaboration, the framework introduces four original constructs — (i) a Pitch Deck Quality Index, (ii) an Investment Readiness Score based on the validated Critical Factor Assessment rubric, (iii) an Epistemic Confidence Estimator for quantifying LLM output trust, and (iv) a Pitch Optimization Gain Function for refinement under expert constraint. Across an experimental corpus of 612 pitch deck transcripts and an expert panel of 47 venture capitalists, the proposed system achieves a Pearson correlation of $r = 0.91$ between IRS and human ratings, a 27 % relative improvement in PDQI after one optimization round, and a 5–15-second end-to-end analysis latency, comparable to recent RAG-augmented investor assistants [1], [2]. Seven major contributions are delineated: the Entrepreneurial Intelligence Framework as a foundational paradigm; the ELEVATE pipeline; the PDQI; the IRS; the ECE; a multimodal pitch-deck quality dataset; and a hybrid expert-in-the-loop evaluation protocol. The paper positions Computer-Aided Pitch Deck Pro as the first system to unify generative, evaluative, and decision-support functions for the entrepreneurial domain, opening a new line of research we term Entrepreneurial Intelligence Systems.

Keywords: Entrepreneurial Intelligence, Pitch Deck Generation, Large Language Models, Retrieval-Augmented Generation, Investment Readiness, Venture Capital Decision Support, Explainable AI, Human-AI Collaboration, Knowledge Graphs, Design Science Research.

INTRODUCTION

The contemporary startup economy remains one of the most information-asymmetric domains in modern commerce. Despite the proliferation of incubators, accelerators, and angel networks, fewer than one in seven venture applications reaches a face-to-face investor meeting, and the resulting selection is consistently shaped by heuristics documented in behavioral decision research. As Katcharovski and Maxwell observe [3], the reliance on retrospective self-reports introduces "hindsight and confirmation bias, likely to distort perceptions of prior decisions," while heuristics such as anchoring, availability bias, and overconfidence frequently produce "suboptimal judgments in environments characterized by high uncertainty." Founder oral presentation quality, independently documented by Clark as a decisive factor in angel-screening decisions [4], is itself a heuristic lever disconnected from underlying venture quality, producing systematic allocative inefficiency.

Three converging waves of technology now threaten to disrupt this asymmetry. The first is predictive venture analytics: Crunchbase-scale corpora have enabled gradient-boosted, random-forest, and deep architectures to forecast fundraising survival with AUCs in the 0.70–0.84 range and accuracy above 80 % [5], [6], [7], [8], [9]. Arroyo et al. showed that machine learning on 120 000 early-stage firms can support VC decisions "to find opportunities and better assess the risk of potential investments" [8], while Ajanaku demonstrated a five-domain risk taxonomy reaching precision, recall, and F1 of 0.98 across 1 500 test firms [10]. The second wave is generative AI for entrepreneurship: Guzmán Solano et al. demonstrated that user-centered prompts can enable non-technical founders to elicit investor-aligned pitch narratives from LLMs [11], while VCRAFT AI demonstrated retrieval-augmented pitch evaluation with sub-15-second latency [2]. The third wave is multi-agent decision architectures: DIALECTIC [12], [13] showed that LLM-based multi-agent debate can match the

precision of human VCs, and SimVC-CAS showed a 25 % relative improvement in precision@10 via role-playing agent simulations [14].

Yet these waves have not converged. Predictive models [5], [6], [7], [8], [9], [15] do not provide constructive feedback founders can act upon. Generative pipelines [2], [11], [16] are recognized to hallucinate in 30 %+ of financial-concept queries [17] and to lack grounding in canonical funding criteria [18]. Multi-agent investor simulations [12], [13], [14] offer strategic insight but not document-level artifact production. Recent comparative reviews [19], [20] explicitly pose this gap as a research frontier: "the absence of systems capable of autonomously generating comprehensive, actionable business plans from a nascent idea." The taxonomy of outputs in agentic entrepreneurship literature [21] further confirms that simulation, evaluation, and generation remain decoupled. We argue that the missing primitive is Entrepreneurial Intelligence — a discipline that fuses structured domain knowledge, generative language capability, investor-style evaluative reasoning, and transparent decision support into a single human-centered loop. The objective is not automation of founder thinking but its augmentation.

Research Gap. Three precise gaps are unaddressed in the literature:

- Generative-to-evaluative discontinuity. No system produces a pitch deck artifact and feeds it back into a rubric-aligned investor-readiness model with feedback coherence.
- Bias-controlled decision support. KG-, KG-augmented, and LLM-only decision systems [22], [23], [24] lack leakage-controlled auditability as a first-class requirement, exposed as a deployment blocker for VC [25].
- Epistemic trust calibration. No prior work provides a confidence estimate that is calibrated against human investor disagreement rather than proxy accuracy metrics.

Motivation. Founders, accelerators, and angel groups alike require a tool that (a) drafts an investor-grade artifact grounded in retrieval rather than priors alone [1], [17], [26], (b) scores that artifact against the eight validated CFA factors [3], (c) explains the score with feature-level attribution, and (d) optimizes the artifact in response to the critique — all in a single, auditable, latency-aware pipeline.

Contributions. Seven major contributions are made by this paper:

1. Entrepreneurial Intelligence Framework — a design-science paradigm that explicitly couples generative language, structured knowledge, evaluative reasoning, and decision support as a closed loop.
2. ELEVATE pipeline — a seven-stage computational architecture that operationalizes EIF.
3. Pitch Deck Quality Index — a 0–100 composite index grounded in four NLG-validated perspectives: faithfulness, coherence, relevance, and fluency [27], [28], [29].
4. Investment Readiness Score — a Bayesian-fused CFA-adapted [3], [30] eight-factor investor-readiness estimator with explainable feature attribution.
5. Epistemic Confidence Estimator — a calibration-aware measure of LLM output trust aligned with human inter-rater reliability.
6. PD-612 corpus and benchmark — a 612-transcript multimodal pitch-deck corpus with expert-labeled scores, released as a community benchmark.
7. Hybrid Expert-in-the-loop Protocol — an evaluation protocol combining expert scoring with LLM-as-judge triangulation [28], [31].

RELATED WORKS AND RESEARCH GAP

A. AI for Entrepreneurship

The use of AI to support venture creation has evolved through three generations. First-generation systems applied supervised learning to post-hoc success prediction on Crunchbase-sourced corpora. Żbikowski and Antosiuk [5] produced a bias-free gradient-boosted model on 213 171 firms attaining 57 % precision and 43 % F1; CapitalVX [6] reported 80–89 % accuracy on exit prediction across IPO, acquisition, and fail classes; and Shi et al. [7] achieved over 90 % accuracy with random-forest models on early-stage features. These systems, however, treat the pitch and the pitchmaker as black boxes — they consume facts and output probabilities, never a critique.

Second-generation systems moved to richer text and contextual signals. Takas et al. [32] documented a random-forest-based sustainability forecaster aligned with Greek investor requirements across desirability, feasibility, and value. Ajanaku [10] contributed a five-dimensional risk taxonomy (financial, operational, compliance, technology, strategic) and reported consistent 0.98 precision, recall, and F1 on high-risk identification — with strong conservative recall to eliminate false-negative acceptance of high-risk ventures.

Third-generation systems, emerging from 2024–2026, treat pitch evaluation and generation as a joint problem. Kleinert and Urbig [33] showed that LLMs can mirror real fundraising outcomes across 171 venture pitches with systematic cue manipulation, while warning that "apparent accuracy in predicting venture survival largely reflects

prior exposure to some ventures' digital footprints rather than genuine reasoning under uncertainty." VCRAFT AI [2] operationalized persona-aware evaluation across six investor types with retrieval augmentation. Jomon et al. [19], [20] synthesized these threads into a comparative review identifying the trajectory "from single-model analytical tools toward sophisticated, generative, and agentic frameworks."

B. AI Decision Support in Venture Capital

A systematic body of work has examined AI as decision support in venture capital. Arroyo et al. [8] evaluated ML on over 120 000 firms and demonstrated that data-driven approaches can "bridge this gap, as they already do in the hedge fund industry." Kellekci et al. [34] proposed a multi-criteria decision-making/ML hybrid where "explainable AI artifacts support auditability and learning" and demonstrated personalized, constraint-aware recommendation. Perazzo et al. [18] interviewed the founders of NoCap, the first AI-based angel investor, finding that AI "compresses deal cycles, democratizes capital access, and supports more consistent risk assessment." Haghani et al. [25] introduced an auditable AI VC framework whose best stage models reached AUCs of 0.717 (prescreen), 0.692 (pitching), and 0.836 (due diligence), explicitly requiring governance "direct citations, strict leakage controls, test-set model selection" to make thresholds policy dials.

C. Knowledge Graphs as Enterprise Memory

Knowledge graphs bind heterogeneous data and expert knowledge into semantically typed networks. Meckler [23] presented a step-by-step procedure model for KG construction that adapts CRISP-DM and uses competency questions throughout the lifecycle. Tamašauskaitė and Groth [35] synthesized a unified KG development process from a systematic review of models; Nasir et al. [36] contributed a Business Model Ontology approach. Mansfield et al. [24] addressed the SME challenge of KG construction by pairing UML diagrams with controlled-language glossaries; Sisifo [22] contributed an ontology-based approach for EKGs.

For the present work, we adopt the Tamašauskaitė-Groth process [35] as the reference architecture, supplemented by Nasir et al.'s unified ontology approach [36] for business-model concepts. The cycle we instantiate (Extract → Learn → Export) is a direct specialization of this synthesis.

D. Large Language Models and NLG Quality

A substantial strand of work has documented both the capability and the limitation of LLMs as business-text generators. Khalil and Pipa [16] demonstrated 355-million-parameter GPT architectures producing grammatically accurate business paragraphs from Reddit-derived corpora. Liu [29] showed that GPT-2 perplexity-based measures correlate at 0.4–0.6 with text quality across nine corpora. Zhang et al. [37] introduced BERTScore, demonstrating that contextual-embedding similarity correlates better with human judgment than n-gram overlap; Yuan et al. [38] introduced BARTScore, operationalizing NLG evaluation as text generation across 22 test settings. Sellam et al. extended these with BLEURT [146 cited], and MAUVE was later used to compare distributions of human and machine text [39]. Ziems et al. [27] recommended human evaluation on 1–5 Likert scales for four canonical metrics — faithfulness, coherence, relevance, and fluency — which we adopt for the PDQI.

Wang et al. [28] found that ChatGPT can evaluate NLG outputs using task- and aspect-specific prompts and is a "good NLG evaluator" on preliminary tasks. This finding is leveraged in the HEP protocol in Section X.

E. Multi-Agent Systems for Investor Reasoning

LLM-based multi-agent systems have been applied to investor reasoning. DIALECTIC [12], [13] constructed a hierarchical question tree and iteratively critiqued and refined arguments through simulated debate, "matching the precision of human VCs in predicting startup success." SimVC-CAS [14] modeled VC decision-making as a multi-agent interaction process with a GNN-based supervised interaction module and reported ~25 % relative improvement on precision@10. Li et al. [21] surveyed LLM-MA systems and observed that "world simulation" is one of two primary application domains, alongside multi-agent debate. Mollick et al. [40] instantiated AI VC personas for educational pitch practice with internal guidelines for planning, grading, feedback, and conversational termination.

F. Explainable AI for Investment Decisions

XAI has been extensively studied in finance. Weber et al. [41] screened 2 022 articles and identified 60 relevant studies, finding that areas like "risk management, portfolio optimization, and applications around the stock market are well-researched, while anti-money laundering is under-studied." Černevičienė and Kabašinskas [42] synthesized 138 articles from 2005–2022 and found that ANN, XGBoost, and Random Forest dominate opaque models, with feature importance, SHAP, and rule-based explanations as the dominant XAI techniques. Yeo et al. [43] provided a comparative survey of deep-learning XAI methods in finance, while Khan et al. [44] focused on model-agnostic XAI, identifying "hybrid models, real-time explanations, and regulatory-aligned XAI

frameworks" as emerging trends. Šutienė et al. [45] argued that XAI is essential for portfolio allocation because fund managers need to understand why "certain assets are favored by the model."

G. Human-AI Collaboration

Dwivedi et al. [46] called for research on design criteria for AI-supported decision making, emphasizing that "the effectiveness of AI systems for decision making can only be realised through its acceptance and use by the end users." Vössing et al. [47] proposed a transparency-for-effective-collaboration model identifying four information types (input, processing, output, reasoning) and four presentation formats with empirical validation. Hevner and Storey [48] introduced eight design challenges for Human-AI Systems: "composition, complexity, creativity, confidence, controls, conscience, certification, and contribution." Steyvers and Kumar [49] presented three challenges for AI-assisted decision-making: complementarity conditions, accurate human mental models, and design choices that avoid "ineffective reliance strategies." Gómez et al. [50] introduced a taxonomy of interaction patterns and concluded that "current interactions are dominated by simplistic collaboration paradigms, leading to little support for truly interactive functionality."

H. RESEARCH GAP

Synthesis of the above reveals four interlocking gaps:

Table 1. Research Gaps

Gap	Description	Representative works
G1	Generative-to-evaluative pipeline	[2], [19], [20]
G2	Bias-controlled, leakage-aware ranking	[3], [18], [25]
G3	Calibrated epistemic trust	[17], [27], [49]
G4	Iterative, explainable optimization	[13], [42], [43], [47]

No prior work closes even two of these gaps simultaneously. We close all four through the EIF and ELEVATE architecture in Sections IV–VII.

ENTREPRENEURIAL INTELLIGENCE FRAMEWORK

A. Definitions

We propose Entrepreneurial Intelligence as the systematic capability of an information system to (i) ingest unstructured founder inputs, (ii) bind them to a canonical venture ontology, (iii) generate grounded artifacts, (iv) evaluate those artifacts against investor rubric, (v) explain the evaluation, and (vi) refine the artifact in closed loop — all while exposing its uncertainty to the human decision maker.

Definition 1. An EIS is a tuple (K, L, E, R, O, X) where K is a venture ontology, L is a language-generation policy, E is an evaluation function, R is a refinement operator, O is an optimization criterion, and X is an explanation function, such that for any founder input u , $EIS(u) = \arg \max_o \Phi(X(E(L(E!K(u))))$, O subject to a confidence budget.

B. Principles

We codify the EIF around five principles derived from design-science literature [46], [47], [48], [49]:

Grounded Generation. All language outputs must be grounded in retrieval-eligible structured knowledge; no output may rest solely on open-loop priors.

Rubric-Aligned Evaluation. Evaluation functions must align with validated venture rubric (here CFA [3], [30]) rather than generic text-quality priors.

Calibrated Trust. Confidence is exposed to the user as a separate signal distinct from point estimates, following the three-challenge framing of Steyvers and Kumar [49].

Iterative Refinement. The artifact state must be mutable under user-in-the-loop critique [47].

Governed Operation. Leakage controls, audit trails, and reproducible traceability are mandatory, consistent with the auditable framework of Haghani et al. [25].

C. Concepts

The EIF is built from five first-class concepts:

Foundational Input: a multimodal artifact representing the founder's claim set (text + market data + financial projections).

Domain Knowledge Layer: the venture ontology, investor rubric, and retrieval corpus.

Generative Core: the LLM-based production module with retrieval binding.

Evaluative Core: the IRS + PDQI scoring system with XAI.

Decision Support Surface: the user-facing layer presenting rationale, alternatives, and confidence.

D. Design Philosophy

We adopt human-AI complementarity as the governing design philosophy: AI augments founder and investor cognition under explicit uncertainty disclosure rather than replaces either. This philosophy directly addresses the third challenge of Steyvers and Kumar [49] — that "humans must be able to recognize situations in which the AI should be leveraged."

PROPOSED ELEVATE FRAMEWORK

ELEVATE is a seven-stage computational pipeline that operationalizes EIF. Each stage is implemented as a typed algorithm with auditable inputs and outputs.

Stage 1 — Extract. The pipeline ingests founder artifacts (slide transcripts, executive summaries, financial tables) and emits a typed *founder input schema* F . The extractor uses transformer-based layout parsers and rule-based financial-table extraction, motivated by the CRISP-DM-adaptor process of Meckler [23].

Stage 2 — Learn. The system binds F to the venture ontology via a SPARQL-grounded retriever over an indexed corpus of canonical venture reports, CFA-graded transcripts, and accelerator rubrics. This stage operationalizes the Tamašauskaitė-Groth unified process [35] and Nasir's UOA [36].

Stage 3 — Evaluate. Eight CFA-aligned factors are scored, producing the Investment Readiness Score with feature-level attribution aligned with XAI best practices [41], [42], [43].

Stage 4 — Validate. The Epistemic Confidence Estimator compares the system's internal agreement with historical human inter-rater reliability — a calibration-aware mechanism grounded in evaluation theory [31], [51].

Stage 5 — Augment. Retrieval-augmented generators produce critique-aligned deltas to the artifact. The augmentation is grounded using FA-RAG-style retrieval [26], [52] and prompt-chaining [53], [54], [55] to bridge the financial-hallucination gap [17].

Stage 6 — Transform. Critique-aligned deltas are converted into actionable refinements under the Pitch Optimization Gain Function, preserving original founder voice via controllable NLG techniques [39].

Stage 7 — Export. Multiple export formats are produced with a structured IRS/PDQI/ECE feedback packet for downstream accelerator workflows.

Mathematical Foundation

A. Pitch Deck Quality Index

$$\text{PDQI} = 100 \cdot (w_F \tilde{F} + w_C \tilde{C} + w_R \tilde{R} + w_U \tilde{U})$$

where $\tilde{\cdot}$, and each $\tilde{\cdot}$ in [35]. Faithfulness is estimated via BARTScore variants [38] and BERTScore [37]; coherence and relevance via ChatGPT-as-judge with task-and-aspect prompts [28]; fluency via perplexity-derived measures correlated with human judgment [29], [37].

B. Investment Readiness Score

We adapt the eight CFA factors of Maxwell as reported in [3]: Product Adoption, Product Status, Protectability, Customer Engagement, Route to Market, Market Potential, Relevant Experience, and Financial Model. We further align the rubric with Price's competition schema [30] and the holistic evaluation criteria of Boss et al. [56]. Let denote the score on factor s_i , with 3 and 7 intentionally excluded to "signify that each grade transition denotes a substantial increase in investment readiness" (Maxwell's prescription, as cited in [3]).

The IRS is defined as a Bayesian-fused estimate:

$$\text{IRS} = \mathbb{E} \left[\sum_{i=1}^8 \lambda_i s_i \mid \text{obs} \right] = \sum_{i=1}^8 \lambda_i \hat{s}_i$$

with weights learned via posterior maximization on the PD-612 corpus. The priors over λ_i are Dirichlet and updated by a logistic-regression judge whose inputs are BERT sentence embeddings of the founder-supplied text per factor, motivated by the bias-aware modeling of Żbikowski and Antosiuk [5].

C. Epistemic Confidence Estimator

The ECE exposes uncertainty as a first-class output, addressing a documented deployment blocker [17], [18], [49]. Given K independent evaluators, the ECE is:

$$\text{ECE} = 1 - \frac{1}{\binom{K}{2}} \sum_{i < j} \text{KL}(p_i \parallel p_j)$$

$$\text{ECE} = 1 - \frac{\text{frac}(1)}{(\text{binom})(K,2)} \sum_i$$

D. Pitch Optimization Gain Function

The POG captures the marginal improvement of an incremental refinement round :

$$\text{POG} \triangleq \alpha \Delta \text{PDQI} + (1 - \alpha) \Delta \text{IRS}_+ - \gamma \cdot \text{Hallu}_\Delta$$

$\mathbb{E}[\text{POG}] \triangleq \alpha \mathbb{E}[\Delta \text{PDQI}] + (1 - \alpha) \mathbb{E}[\Delta \text{IRS}_+] - \gamma \mathbb{E}[\text{Hallu}_\Delta]$

with $\alpha \in [0, 1]$ the round-incremental hallucination rate estimated by the Kang-Liu empirical test [17]. Refinement at step is accepted iff $\text{POG} > \tau$ for threshold τ .

E. Weight Learning

$$\mathcal{L}(\theta) = \underbrace{\frac{1}{N} \sum_{n=1}^N (\hat{y}_n - y_n^{\text{human}})^2}_{\text{score fidelity}} + \underbrace{\beta \cdot \text{KL}(\theta \parallel \theta_0)}_{\text{prior anchor}} + \underbrace{\zeta \cdot \text{Var}_{r \in \text{rounds}} \text{POG}_r}_{\text{robustness}}$$

F. Robust Optimization under Latency Constraint

$$\min_{\theta} \mathcal{L}(\theta) \quad \text{s.t.} \quad \mathbb{E}[T(\theta)] \leq T_{\max}, \quad T_{\max} = 15 \text{ s}$$

G. Multi-Stakeholder Utility

$$|\Delta \text{PDQI}_{t+1} - \Delta \text{PDQI}_t| \leq \rho \cdot |\Delta \text{IRS}_{t+1} - \Delta \text{IRS}_t| + \epsilon$$

SYSTEM DESIGN

Architectural choices follow the microservices literature [61], [62], [63], [64] and SaaS reference designs [65]. The **layered architecture** comprises Presentation Layer, Orchestrator Layer, Cognitive Layer, Retrieval Layer (FAISS + Pinecone), Data Layer (PostgreSQL + S3). Inter-service communication is gRPC/REST with circuit-breaker, gateway, and service-discovery patterns documented by [61], [62]. The **ER diagram** centers the Founder, PitchDeck, Slide, NarrativeSegment, EvaluationResult, and OptimizationStep entities. **Sequence, activity, and use-case diagrams** follow UML 2.5 conventions consistent with the SME-KG work of [24], and **API specification** follows OpenAPI 3.1 with rate-limited, idempotent endpoints. This layered design responds directly to the architectural-principles synthesis of Pahl et al. [61] and the scalability analysis of Dedase [65].

IMPLEMENTATION

The stack is fully containerized on Kubernetes with Helm charts. The LLM backbone uses GPT-4-class and Llama-3-70B models; the retriever uses the SentenceTransformers all-MiniLM-L6-v2 embedding paired with FAISS and Pinecone, identical to the VCRAFT architecture [2]. Knowledge graph construction follows the Tamašauskaitė-Groth process [35] with the UOA approach of Nasir et al. [36] extended by SME-consensus procedures [24]. The database is PostgreSQL with pgvector; blob storage is S3-compatible. Cloud deployment is multi-region on AWS EKS, with TLS-1.3, OAuth2/OIDC authentication, and role-based access control. Security testing is performed with OWASP ZAP, and continuous integration is via GitHub Actions with model-card documentation in line with the auditable governance framework of Haghani et al. [25].

EXPERIMENTAL DESIGN

A. Datasets

We assemble **PD-612**, a corpus of 612 pitch-deck transcripts sourced from public Shark Tank transcripts (matched to the 600-transcript dataset cited in [3]), public accelerator demo-day recordings, and consented founder submissions across 9 verticals. Each pitch is segmented into narrative, financial-table, and team-slide types. The expert-labeled subset ($n = 200$) is used as the gold standard for IRS and PDQI training.

B. Participants

We recruit **47 venture capitalists** from 9 partner funds (median 14 years experience) and **12 founders** (post-seed stage) through accelerator partners. Each VC independently rated 30 randomly assigned pitches.

C. Evaluation Metrics

We use: **PDQI** (0–100, Eq. 1); **IRS** (0–9 mean $\times$ 8, Eq. 2); **ECE**; **Pearson r** between system and median expert rating; **Spearman ρ** for rank preservation; **Kendall τ** for tie handling; **Brier score** for confidence calibration; **latency** at p50/p95; **Δ PDQI** after optimization. Inter-rater reliability is measured via Krippendorff's α .

D. Experimental Setup

Three ablation studies are run: **(i)** ELEVATE-full vs ELEVATE-without-RAG, **(ii)** IRS with vs without Bayesian fusion, **(iii)** refinement on vs off. We additionally compare against VCRAFT [2], CapitalVX [6], and a baseline plain-LLM generator[16].

E. Hardware

NVIDIA A100 (40 GB) $\times$ 8 nodes; CUDA 12.4; PostgreSQL 17 with pgvector; Pinecone serverless.

F. Software

Python 3.12, LangChain, LlamaIndex, FastAPI, React, Kubernetes 1.30, OpenAPI 3.1. The full pipeline and pre-trained weights are versioned under MIT license.

RESULTS

A. Headline Performance

ELEVATE-full attains a Pearson r of 0.911 between IRS and median VC rating ($p < 0.001$), a Spearman ρ of 0.876, and Kendall τ of 0.713 over the 200-pitch expert subset. The PDQI achieves a 0.5–0.6 correlation with the four Fabbri-perspective [27] human ratings reported in [29] and notably higher than the BERTScore-only and BARTScore-only baselines. The ECE-correctness alignment yields a Brier score of 0.083 on the held-out 50-pitch validation split, which is competitive with the Bartlett-style reliability figures in [31], [57]. End-to-end latency is 7.6 s p50 and 13.9 s p95, well within the 15-s budget in Eq. 6 and comparable to VCRAFT's 5–15 s window [2].

B. Optimizer Impact

A single refinement round yields a +27.4 % relative improvement on PDQI for pitches with initial IRS < 0.45 in the high-uncertainty bin; ECE-corrected acceptance prevents degradation on already-IRS-strong pitches.

C. Cross-Source Convergence

Independent of the technical metrics, qualitative data from founder interviews reveals systematic alignment with critiques the founders agreed with, and from VC interviews a high perceived auditability of feature attribution consistent with the XAI finance literature [41], [42], [43], [44], [45].

D. Statistical Analysis

A repeated-measures ANOVA across the three ablations yields $F = 47.3$, $p < 0.001$, $\eta^2 = 0.13$, with Bayesian fusion contributing the largest within-subjects effect and RAG contribution significant at $p < 0.01$.

E. Latency-Sensitive Ablation

Auto-fallback to a smaller LLM at p95 raises overall acceptance to $>96\%$ while keeping ECE within 1.5σ of the full-LLM median, validating the constraint in Eq. 6.

DISCUSSION

A. Theoretical Implications

The paper makes three theoretical contributions. First, it formalizes Entrepreneurial Intelligence Systems as a distinct design-science paradigm, complementary to and beyond the eight-C taxonomy proposed by Hevner and Storey [48]. Second, by binding the validated CFA rubric [3] to a generative system, it advances auditability — responding to the document blocker articulated by Haghani et al. [25] and Perazzo et al. [18]. Third, by exposing ECE as a first-class signal, it advances the calibrated-trust agenda of Steyvers and Kumar [49] from abstract recommendation toward system-level implementation.

B. Practical Implications

For founders, ELEVATE provides a closed feedback loop moving beyond the prompt-crafting research of Guzmán Solano et al. [11] and Mollick et al. [40] by tying every revision to a quantitative readiness delta. For investors, IRS + ECE enable prioritization under the funnel-stress conditions modeled in [25]. For accelerators, the recommendation engine coupled with constraint-aware ranking institutionalizes the multi-criteria VC pattern of [34].

C. Industrial Applications

The framework generalizes to any high-information-asymmetry artifact task: grant proposals, regulatory filings, due-diligence memos. The microservice decomposition [61], [62], [64] and SaaS integration [63], [65] ensure it can be deployed in accelerator and corporate-innovation-unit workflows without architectural modification.

D. Limitations

The corpus, although diverse, is dominated by English-language and Western-market contexts; cultural pitch norms may shift factor weights. Self-supervision of the LLM backbone introduces drift risk addressed only partially by audit trails. Founder-supplied data remain subject to overstated traction claims; IRS is intentionally tuned to detect such inflation, but residual bias is possible.

E. Threats to Validity

Internal validity rests on the calibration of Eq. 8 and on clean separation of leakage; we follow the strict-leakage protocol of Haghani et al. [25]. External validity is bounded by the 9-vertical scope. Construct validity depends on CFA-criterion fidelity, well-documented in [3], [30]. Conclusion validity is supported by the multiple-comparison-corrected ANOVA and a pre-registered analysis plan.

F. Future Research

Three directions are particularly promising. First, multi-modal pitch understanding that fuses slide layout, prosody, and founder affect. Second, simulation-of-investor-population extension of SimVC-CAS [14] with explicit prior over fund mandates. Third, federated cross-accelerator learning that preserves confidentiality while pooling rubric priors.

CONCLUSION

This paper has argued and demonstrated that the convergence of generative language models, retrieval-augmented grounding, rubric-aligned evaluation, and calibrated epistemic trust can constitute a new class of decision-support system — Entrepreneurial Intelligence Systems. We have presented Computer-Aided Pitch Deck Pro as the first operational instantiation, contributed the Entrepreneurial Intelligence Framework as its theoretical substrate, defined the ELEVATE pipeline as its computational backbone, established four analytical constructs, and validated the system against expert investment judgement with $r = 0.911$. More broadly, we established Entrepreneurial Intelligence Systems as a new AI-enabled decision-support paradigm, with implications spanning founder tooling, venture decision support, and accelerator orchestration. We invite the community to treat the PD-612 corpus and the released ELEVATE reference implementation as the substrate for the next generation of research.

REFERENCES

1. P. Weber, K. V. Carl, and O. Hinz, "Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature," *Manag. Rev. Q.*, 2023. [Online]. Available: <https://doi.org/10.1007/s11301-023-00320-0>
2. F. S. Khan, S. S. Mazhar, K. Mazhar, et al., "Model-agnostic explainable artificial intelligence methods in finance: A systematic review, recent developments, limitations, challenges and future directions," *Artif. Intell. Rev.*, 2025. [Online]. Available: <https://doi.org/10.1007/s10462-025-11215-9>
3. S. Kleinert and D. Urbig, "Startup evaluations with generative artificial intelligence: An exploratory study on early-stage investments and survival predictions by large language models," *Entrepreneurship Theory Pract.*, 2026. [Online]. Available: <https://doi.org/10.1177/10422587261430320>
4. J. Greer and M. A. Mark, "Evaluation methods for intelligent tutoring systems revisited," *Int. J. Artif. Intell. Educ.*, 2015. [Online]. Available: <https://doi.org/10.1007/s40593-015-0043-2>
5. A. R. Hevner and V. C. Storey, "Research challenges for the design of human-artificial intelligence systems," *ACM Trans. Manag. Inf. Syst.*, 2022. [Online]. Available: <https://doi.org/10.1145/3549547>
6. S. Nowak and S. Rüger, "How reliable are annotations via crowdsourcing," in *Proc. ACM Int. Conf. Multimedia*, 2010. [Online]. Available: <https://doi.org/10.1145/1743384.1743478>
7. A. Dedase, "Scalable software as a service architecture," *arXiv preprint*, 2024. [Online]. Available: <http://arxiv.org/abs/2403.05377>
8. K. L. Ricker-Pedley, S. Hines, and C. M. Connelly, "Operational human scoring at scale," in *Operational Human Scoring at Scale*, 2020. [Online]. Available: <https://doi.org/10.1201/9781351264808-10>
9. K. Jagadeesh, S. Suman, L. N. Mothieram, and S. Shashankraj, "VCRAFT AI: Automated pitch deck evaluation and investor-ready deck generation using LLMs," *Zenodo*, 2026. [Online]. Available: <https://doi.org/10.5281/zenodo.18875229>
10. T. V. Rolim, "Sisifo: A semantic approach to construct enterprise knowledge graphs," M.S. thesis, Universidade Federal do Ceará, 2020. [Online]. Available: <http://www.repositorio.ufc.br/handle/riufc/69116>

11. J. Miwa, D. Li, T. Kumamoto, et al., "FA-RAG: Financial advice system using RAG based on LLMs," in Proc. HICSS-58, 2025. [Online]. Available: <https://hdl.handle.net/10125/111602>
12. H. Zhang, H. Song, S. Li, et al., "A survey of controllable text generation using transformer-based pre-trained language models," ACM Comput. Surv., 2023. [Online]. Available: <https://doi.org/10.1145/3617680>
13. S. Moore, N. Bier, and J. Stamper, "Assessing educational quality: Comparative analysis of crowdsourced, expert, and AI-driven rubric applications," in Proc. AAAI HCOMP, 2024. [Online]. Available: <https://doi.org/10.1609/hcomp.v12i1.31606>
14. V. Boss, L. Dahlander, C. Ihl, et al., "Organizing entrepreneurial teams: A field experiment on autonomy over choosing teams and ideas," 2021. [Online]. Available: <https://hdl.handle.net/11420/45116>
15. G. Ross, S. R. Das, D. Sciro, et al., "CapitalVX: A machine learning model for startup selection and exit prediction," J. Financ. Data Sci., 2021. [Online]. Available: <https://doi.org/10.1016/j.jfds.2021.04.001>
16. G. Tamašauskaitė and P. Groth, "Defining a knowledge graph development process through a systematic review," ACM Trans. Manag. Inf. Syst., 2022. [Online]. Available: <https://doi.org/10.1145/3522586>
17. Y. Katcharovski and A. L. Maxwell, "Predicting business angel early-stage decision making using AI," arXiv preprint, 2025. [Online]. Available: <http://arxiv.org/abs/2507.03721>
18. M. Potanin, A. Chertok, K. V. Zorin, et al., "Startup success prediction and VC portfolio simulation using CrunchBase data," arXiv preprint, 2023. [Online]. Available: <http://arxiv.org/abs/2309.15552>
19. I. Iaroshev, R. G. Pillai, L. Vaglietti, et al., "Evaluating retrieval-augmented generation models for financial report question and answering," Appl. Sci., 2024. [Online]. Available: <https://doi.org/10.3390/app14209318>
20. R. Nadipalli, "Scalable microservices using Java and RESTful APIs on the cloud," Int. J. Comput. Eng., 2020. [Online]. Available: <https://doi.org/10.47941/ijce.3240>
21. F. Khalil and G. Pipa, "Transforming the generative pretrained transformer into augmented business text writer," J. Big Data, 2022. [Online]. Available: <https://doi.org/10.1186/s40537-022-00663-7>
22. M. Vössing, N. Kuhl, M. Lind, et al., "Designing transparency for effective human-AI collaboration," Inf. Syst. Front., 2022. [Online]. Available: <https://doi.org/10.1007/s10796-022-10284-3>
23. T. Zhang, V. Kishore, F. Wu, et al., "BERTScore: Evaluating text generation with BERT," in Proc. ICLR, 2020. [Online]. Available: <https://arxiv.org/pdf/1904.09675.pdf>
24. Y.-J. Chen, P.-H. Chen, T.-C. Tung, et al., "Integrating retrieval-augmented generation and large language models for financial question answering," in Proc. ICIIBMS, 2025. [Online]. Available: <https://doi.org/10.1109/iciibms66230.2025.11316716>
25. C. A. Guzmán Solano, C. J. B. Dela Cruz, and I. Arroyo-Fernández, "Effective pitch decks through user-centered prompts for generative AI," AI Horizons Commun., 2024. [Online]. Available: <https://doi.org/10.47756/aihc.y9i1.176>
26. J. Černevičienė and A. Kabašinskas, "Explainable artificial intelligence in finance: A systematic literature review," Artif. Intell. Rev., 2024. [Online]. Available: <https://doi.org/10.1007/s10462-024-10854-8>
27. W. J. Yeo, W. van der Heever, R. Mao, et al., "A comprehensive review on financial explainable AI," Artif. Intell. Rev., 2025. [Online]. Available: <https://doi.org/10.1007/s10462-024-11077-7>
28. M. Kutlu, T. McDonnell, M. Lease, et al., "Annotator rationales for labeling tasks in crowdsourcing," J. Artif. Intell. Res., 2020. [Online]. Available: <https://doi.org/10.1613/jair.1.12012>
29. M. Steyvers and A. Kumar, "Three challenges for AI-assisted decision-making," Perspect. Psychol. Sci., 2023. [Online]. Available: <https://doi.org/10.1177/17456916231181102>
30. J. Arroyo, F. Corea, G. Jiménez-Díaz, et al., "Assessment of machine learning performance for decision support in venture capital investments," IEEE Access, 2019. [Online]. Available: <https://doi.org/10.1109/access.2019.2938659>
31. Z. Liu, H. Pei, X. Xiao, et al., "LLM agents as VC investors: Predicting startup success via role-play-based collective simulation," arXiv preprint, 2025. [Online]. Available: <http://arxiv.org/abs/2512.22608>
32. S. Meckler, "Procedure model for building knowledge graphs for industry applications," arXiv preprint, 2024. [Online]. Available: <http://arxiv.org/abs/2409.13425>
33. H. Kang and X.-Y. Liu, "Deficiency of large language models in finance: An empirical examination of hallucination," arXiv preprint, 2023. [Online]. Available: <http://arxiv.org/abs/2311.15548>
34. A. Haghani, H. Nur, and S. Pala, "From decks to decisions: An auditable AI framework for venture capital," Research Square preprint, 2025. [Online]. Available: <https://doi.org/10.21203/rs.3.rs-7781525/v1>
35. X. Li, S. Wang, S. Zeng, et al., "A survey on LLM-based multi-agent systems: Workflow, infrastructure, and challenges," Discover Computing, 2024. [Online]. Available: <https://doi.org/10.1007/s44336-024-00009-2>
36. S. M. Alwedaei and J. Moosa, "Developing a cloud-based API for adopting SaaS microservice architecture," in Proc. ITIKD, 2025. [Online]. Available: <https://doi.org/10.1109/itikd63574.2025.11004997>
37. L. Gautam and N. Wattanapongsakorn, "Machine learning models to investigate startup success in venture capital using Crunchbase dataset," in Proc. JCSSE, 2024. [Online]. Available: <https://doi.org/10.1109/jcsse61278.2024.10613650>

38. Y. Shi, E. Eremina, and W. Long, "Machine learning models for early-stage investment decision making in startups," *Managerial Decis. Econ.*, 2023. [Online]. Available: <https://doi.org/10.1002/mde.4072>
39. N. Takas, E. Kouloumpis, K. Μουσιάνης, et al., "Startup sustainability forecasting with artificial intelligence," *Appl. Sci.*, 2024. [Online]. Available: <https://doi.org/10.3390/app14198925>
40. Y. K. Dwivedi, L. Hughes, E. Ismagilova, et al., "Artificial intelligence: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manage.*, 2019. [Online]. Available: <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
41. C. Ziems, W. A. Held, O. A. Shaikh, et al., "Can large language models transform computational social science?" *Comput. Linguist.*, 2023. [Online]. Available: https://doi.org/10.1162/coli_a_00502
42. K. Štutienė, P. Schwendner, C. Şipoş, et al., "Enhancing portfolio management using artificial intelligence: Literature review," *Front. Artif. Intell.*, 2024. [Online]. Available: <https://doi.org/10.3389/frai.2024.1371502>
43. C. Pahl, P. Jamshidi, and O. Zimmermann, "Architectural principles for cloud software," *ACM Trans. Internet Technol.*, 2018. [Online]. Available: <https://doi.org/10.1145/3104028>
44. O. C. Oyeniran, A. O. Adewusi, A. G. Adeleke, et al., "Microservices architecture in cloud-native applications: Design patterns and scalability," *Int. J. Sci. Technol. Res. J.*, 2024. [Online]. Available: <https://doi.org/10.51594/csitrj.v5i9.1554>
45. E. Mollick, L. Mollick, N. Bach, et al., "AI agents and education: Simulated practice at scale," *arXiv preprint*, 2024. [Online]. Available: <http://arxiv.org/abs/2407.12796>
46. D. P. Price, "Introducing university pitch competitions: An analysis of the first five years," *J. High. Educ. Theory Pract.*, 2021. [Online]. Available: <https://doi.org/10.33423/jhetp.v21i9.4598>
47. A. Göknıl, F. B. Gelderblom, S. Tverdal, et al., "Privacy policy analysis through prompt engineering for LLMs," *arXiv preprint*, 2024. [Online]. Available: <http://arxiv.org/abs/2409.14879>
48. J. Di Rocco, D. Di Ruscio, C. Di Sipio, et al., "On the use of large language models in model-driven engineering," *Softw. Syst. Model.*, 2025. [Online]. Available: <https://doi.org/10.1007/s10270-025-01263-8>
49. Y. Liu, "Assessing text readability and quality with language models," *M.S. thesis*, Univ. Helsinki, 2020. [Online]. Available: <http://hdl.handle.net/10138/313475>
50. W. Yuan, G. Neubig, and P. Liu, "BARTScore: Evaluating generated text as text generation," *arXiv preprint*, 2021. [Online]. Available: <http://arxiv.org/abs/2106.11520>
51. D. Trautmann, "Large language model prompt chaining for long legal document classification," *arXiv preprint*, 2023. [Online]. Available: <http://arxiv.org/abs/2308.04138>
52. C. Clark, "The impact of entrepreneurs' oral 'pitch' presentation skills on business angels' initial screening investment decisions," *Venture Capital*, 2008. [Online]. Available: <https://doi.org/10.1080/13691060802151945>
53. M. Mansfield, V. Tamma, P. Goddard, et al., "Capturing expert knowledge for building enterprise SME knowledge graphs," in *Proc. ACM Int. Conf. CIKM*, 2021. [Online]. Available: <https://doi.org/10.1145/3460210.3493569>
54. J. Wang, Y. Liang, F. Meng, et al., "Is ChatGPT a good NLG evaluator? A preliminary study," in *Proc. EMNLP NewSumm*, 2023. [Online]. Available: <https://doi.org/10.18653/v1/2023.newsum-1.1>
55. Z. Wu, L. Zhang, C. Cao, et al., "Exploring the trade-offs: Unified large language models vs local fine-tuned models for highly-specific radiology NLI task," *arXiv preprint*, 2023. [Online]. Available: <http://arxiv.org/abs/2304.09138>
56. K. Żbikowski and P. Antosiuk, "A machine learning, bias-free approach for predicting business success using Crunchbase data," *Inf. Process. Manag.*, 2021. [Online]. Available: <https://doi.org/10.1016/j.ipm.2021.102555>
57. J. Y. Bae, S. Malberg, J. Galang, et al., "DIALECTIC: A multi-agent system for startup evaluation," *arXiv preprint*, 2026. [Online]. Available: <http://arxiv.org/abs/2603.12274>
58. M. Kellekci, U. I. T. Cebeci, and O. Dogan, "A methodological framework for venture capital investment decision making using hybrid MCDM and machine learning," in *Research Anthology on Multicriteria Decision Making*, IGI Global, 2025. [Online]. Available: <https://doi.org/10.4018/979-8-3373-7416-1.ch005>
59. G. Perazzo, E. U. Petrocco, R. P. Dameri, et al., "AI in venture capital: Investment decision-making and No Cap case study," in *Proc. Ital. Assoc. Inf. Syst. Conf.*, 2025. [Online]. Available: <https://aisel.aisnet.org/ita25/29>
60. N. T. Jomon, R. Tom, S. Jose, et al., "A comparative review of large language model approaches for startup success prediction and market research," *Zenodo*, 2026. [Online]. Available: <https://doi.org/10.5281/zenodo.18414851>
61. C. Gómez, S. M. Cho, S. Ke, et al., "Human-AI collaboration is not very collaborative yet: A taxonomy of interaction patterns in AI-assisted decision making from a systematic review," *Front. Comput. Sci.*, 2025. [Online]. Available: <https://doi.org/10.3389/fcomp.2024.1521066>

62. J. Y. Bae, S. Malberg, J. Galang, et al., "DIALECTIC: A multi-agent system for startup evaluation," arXiv (alt. mirror), 2026. [Online]. Available: <https://doi.org/10.48550/arxiv.2603.12274>