International Journal of Computer Science and Engineering Development
Volume 1, Issue 4 | September - October 2026 | www.ijcsed.com
ISSN: 3139-0862

A Survey of an Intelligent Framework for Startup Ideation, Business Planning, and Entrepreneurial Growth

Ajay Chopra, Vedanti Patil, Sannidhi Shetty, Sumayya Shaikh, Prof. Sandeep More
Department of Computer Engineering, Watumull Institute of Engineering & Technology, Ulhasnagar, University of Mumbai, Maharashtra, India

***Abstract*—**
This paper surveys ten studies relevant to AI-assisted entrepreneurship and synthesizes their findings into a proposed intelligent framework for startup ideation, business planning and entrepreneurial growth. The reviewed work spans AI adoption, ChatGPT-based idea evaluation, intellectual-property strategy, natural-language processing, startup initiative analysis, sentiment modeling, lean software practices, retrieval-augmented blueprint generation and knowledge-graph-based risk assessment. A qualitative comparison identifies complementary capabilities and persistent concerns involving evidence quality, data availability, privacy, bias and generalizability. Building on these findings, a venture-stage-aware advisory architecture and methodology flowchart are proposed for both aspiring and existing entrepreneurs. The design separates user facts, retrieved evidence and generated assumptions, and supports human validation and iterative revision. This is a literature review and conceptual methodology; no implementation results or startup-success predictions are claimed.

***Keywords*—Artificial intelligence, startup ideation, entrepreneurship, business planning, survey, large language models, retrieval-augmented generation, risk assessment.**

I. INTRODUCTION

Entrepreneurship requires iterative decisions about customer needs, differentiation, resources, market entry and growth. Founders typically combine informal advice, market searches and spreadsheets; this fragmentation can make assumptions difficult to track. Generative artificial intelligence (AI) offers natural-language interfaces for ideation and planning, but its fluent output is not automatically verified business evidence. This survey examines ten studies covering startup evaluation, organizational AI adoption, natural-language processing (NLP), social-media analysis, lean practices, retrievalaugmented generation (RAG) and predictive risk assessment [1]–[10].

The review asks: (RQ1) which AI capabilities are relevant to startup ideation, planning and growth; (RQ2) what methodological and practical limitations recur; and (RQ3) how can these findings inform a responsible, stageaware conceptual architecture? The outcome is a literaturegrounded design proposal, not a claim of implemented or experimentally validated software.

II. SURVEY SCOPE AND ANALYTICAL APPROACH

The review corpus comprises the ten previously selected studies [1]–[10]. They are examined qualitatively according to research objective, technical approach, contribution to entrepreneurial decision support, limitation and relevance to the proposed framework. The set includes peer-reviewed work and one reference supplied as a research manuscript [4]; it is a focused narrative survey rather than a systematic review of all published literature. Claims about individual studies are separated from recommendations derived by synthesis.

III. REVIEW OF THE TEN RESEARCH PAPERS

A. Startups and Artificial Intelligence [1]

Rios-Campos et al. [1] examine the relationship between startups and AI, framing AI as an enabling technology for business processes and innovation while raising ethical and adoption concerns. For entrepreneurial platforms, this supports treating AI as assistance rather than an autonomous authority. Its broad survey perspective does not specify an end-to-end personalized startup-planning workflow or establish performance for a particular implementation.

B. Strategic Integration of Generative AI [2]

Al-Kfairy [2] discusses generative-AI integration in organizational settings, including opportunities, governance, implementation challenges and adoption requirements. The relevance extends beyond newly founded businesses: existing ventures must consider organizational readiness, privacy and responsible use when adopting AI recommendations. Because the emphasis is strategic integration, it does not by itself provide a startup-specific architecture or experimentally validate a unified entrepreneurial assistant.

C. ChatGPT-Based Startup Idea Evaluator [3]

Davalas and Angelaki [3] describe a conversational startup idea evaluator with structured scoring, SWOToriented assessment and funding-related suggestions. This is a direct precedent for turning an open-ended idea into a reviewable entrepreneurial artifact. A key boundary is that an AI-generated evaluation cannot replace actual customer discovery, independently checked market evidence or subsequent operational assessment of an established firm.

D. AI for Startup Growth and IP Strategy [4]

Reddy et al. [4] discuss AI-enabled startup growth and intellectual-property (IP) strategy, including a practical web-

ISSN: 3139-0862

platform orientation. The study broadens the survey beyond ideation to competitive differentiation and protection of intangible assets. Its relevance is conceptual: growth guidance and IP considerations should appear in a broader business workflow. The bibliographic record available in the supplied manuscript citation is incomplete; specific experimental effectiveness should therefore not be inferred.

E. NLP in Model-to-Model Transformations [5]

Danenas and Skersys [5] investigate extracting information and relationships from natural language in model-to-model transformation settings. Although not a startup study, it motivates converting founders' unstructured descriptions into structured entities such as target users, budget, industry and constraints. Transfer to entrepreneurship is an architectural inference; domainspecific extraction accuracy would require separate testing.

F. Startup Initiative Response Analysis (SIRA) [6]

Alotaibi et al. [6] present the SIRA framework for analyzing responses to startup initiatives on Twitter through

text analytics. It shows how public reactions can become signals about reception of an initiative. Such signals may support early market hypotheses, but social-media samples are not representative of every customer segment, and public reactions do not independently establish willingness to pay or business viability.

G. Fuzzy-Sentiment Tweet Analysis [7]

Phan et al. [7] propose a feature-ensemble approach for sentiment analysis of tweets containing ambiguous or fuzzy sentiment. Their contribution highlights the difficulty of interpreting informal online feedback. In an entrepreneurial assistant, this supports cautious sentiment summaries and explicit uncertainty labels; it does not prove that tweet sentiment can forecast startup revenue or product-market fit.

H. Lean Software Startup Practices [8]

Cico [8] studies lean software startup practices in software engineering education, emphasizing experimentation and learning in an iterative environment. This perspective suggests that generated plans should end with testable assumptions, minimum viable product (MVP) activities and feedback cycles. Educational and softwarestartup contexts may not generalize directly to every sector, so the workflow should remain adaptable.

I. RAG-Based Startup Blueprint Generator [9]

Mansuri et al. [9] present Start Genie AI, which combines RAG and large language models (LLMs) to create structured startup blueprints. The reported pipeline uses data preparation, embeddings, vector retrieval and context-aware generation, with a React frontend and FastAPI backend. This work offers a concrete blueprint-generation precedent. Its usefulness remains dependent on the relevance, freshness and credibility of retrieved sources; generating a polished plan does not establish that its business assumptions are true.

J. Multi-Layer Startup Risk Decision Support [10]

Pan, Pan and Sun [10] propose a layered decision-support model combining knowledge graphs, graph convolutional networks, federated learning and deep neural networks to assess startup success and risk using financial and nonfinancial factors. The authors report 93% accuracy and 0.90 AUC in their study; these are their reported results, not results of the framework proposed here. They also acknowledge that a 20-startup experimental sample limits generalization and that binary success/failure categories omit intermediate or pivoting states. The paper motivates risk-factor explanations, but implementing its predictive architecture is outside the present conceptual proposal.

IV. CROSS-STUDY ANALYSIS AND RESEARCH GAP

The ten studies form five complementary strands. Broad adoption and governance research [1], [2], [4] identifies organizational constraints; idea evaluation [3] structures early decisions; NLP and social-signal work [5]–[7] helps interpret text and feedback; lean methods [8] support iterative validation; and blueprint generation and predictive risk research [9], [10] provide more specialized technical exemplars. No single reviewed approach establishes, for this project, a validated conversational workflow covering both aspiring and existing entrepreneurs, with explicit evidence/assumption separation and user-led revision.

Comparative synthesis: [3] is strongest as a precedent for structured idea scoring; [9] for evidence-supported plan drafting; [10] for relational risk modeling; [6], [7] for interpreting public responses; and [8] for iterative learning. The proposed design deliberately does not assume that sentiment analysis equals market validation, that RAG eliminates hallucinations, or that a risk classifier can reliably predict every startup's future.

A. Comparative Classification of the Surveyed Methods

A closer comparison reveals differences in each study's unit of analysis, input evidence and intended output. RiosCampos et al. [1] and Al-Kfairy [2] consider AI adoption at a broad organizational level. The work of Davalas and Angelaki [3] addresses one entrepreneur's proposed idea through a structured evaluation artifact, while the growth and intellectual-property perspective in [4] widens the decision context to an operating venture. Danenas and Skersys [5] consider extraction from natural-language descriptions, a useful but indirect methodological analogue. The SIRA framework [6] and fuzzy-sentiment analysis [7] analyze public textual signals rather than private founder profiles. Lean practice [8] concerns the process of testing assumptions; the RAG-based generator [9] focuses on creating a coherent blueprint; and the layered system of Pan et al. [10] addresses data-intensive risk classification. Accordingly, comparisons based only on the existence of an AI model obscure major differences in tasks, input populations, success criteria and verification requirements.

B. Nature and Reliability of Evidence

The studies also imply an evidence hierarchy. Foundersupplied constraints are useful for personalization but may be incomplete or mistaken. Retrieved documents can support claims about markets, competitors and regulations, yet their provenance and publication dates must be exposed, as suggested by the retrieval-oriented design in [9]. Public reactions collected through methods such as [6] and [7] reveal aspects of discourse, not representative sales intent. Numerical risk indicators such as those studied in [10] may be informative only within the sampling and modeling assumptions of the underlying dataset. An advisory platform

ISSN: 3139-0862

should therefore avoid presenting all sources of information at the same confidence level. A generated output can separately display what the user stated, what a cited source reports, and what the model merely proposes. It should also record which important questions have no adequate evidence. This separation is a design inference from the ten studies, rather than a feature demonstrated by every referenced system.

C. Differences Between New and Established Ventures

For a prospective founder, the central uncertainties concern the identity of the customer, urgency of a problem, candidate value propositions and feasibility with limited resources. The structured scoring perspective in [3] can organize alternative ideas, while lean experiments in [8] motivate subsequent interviews and minimum viable product tests. An existing business instead begins with observed customers, processes and products, but may need to identify bottlenecks, retention problems or expansion opportunities. Adoption and governance concerns discussed in [1] and [2], together with the growth orientation of [4], are therefore more prominent. This distinction supports two related but separate workflows. Both may use profile

extraction [5], evidence retrieval [9] and human review; however, their output questions and validation actions should differ. The novelty of the proposed framework lies in orchestrating these complementary capabilities for two stages, not in claiming that a new model has been invented.

D. Comparative Methodological Limitations

Several recurring limitations prevent direct comparisons of reported results. First, a conceptual or narrative contribution cannot be ranked against a predictive model using accuracy alone. Second, samples of public posts [6], [7] can overrepresent highly active users and omit customers who do not engage online. Third, the context in which lean practices were studied [8] differs from that of non-software enterprises. Fourth, RAG systems [9] depend on an external corpus whose coverage and freshness may vary by industry or location. Fifth, the limited startup sample and binary outcome formulation acknowledged in [10] constrain the interpretation of predictive success metrics. Finally, the publication record for the manuscript [4] remains insufficiently verified to support strong empirical claims. These observations motivate a multicriteria, scenario-based evaluation of the proposed framework rather than a single headline accuracy number.

E. Consolidated Research Gap and Design Implications

The comparison suggests four requirements not jointly established by the reviewed studies: venture-stage recognition, structured yet revisable advice, source-aware uncertainty communication and an evaluation process spanning both idea discovery and business improvement. A system satisfying those requirements would collect a minimal profile, identify missing constraints, generate more than one candidate action, label source status, and preserve user feedback during revision. Domain-specific tasks could call different advisory modules without implying that a sentiment classifier or knowledge-graph risk engine is already implemented. This deliberately modest proposal is consistent with the available literature: the sources motivate architectural components, but they do not collectively validate an integrated deployed platform.

V. PROPOSED METHODOLOGY

A. Design Goals and Venture Profiles

The proposed system is a modular AI-assisted advisory framework for two paths: (1) new venture exploration and (2) existing business improvement. A user provides an objective, business stage, target customers, industry, skills, resources, geography and constraints. Existing ventures may optionally provide customer, revenue-band and operating information. Structured profile extraction is motivated by NLP research [5], while distinct startup stages address the gap identified in the synthesis.

B. Proposed System Architecture

Figure 1 presents a conceptual architecture. A frontend gathers the venture profile; a context manager normalizes constraints; an LLM prompt orchestrator prepares taskspecific requests; and an optional RAG service retrieves source-attributed evidence. Advisory modules produce idea alternatives, SWOT observations, business-model hypotheses, customer-validation tasks, marketing options and a phased roadmap. An evidence layer tags user facts, retrieved claims and model assumptions separately. The

human reviewer can request revisions and export an updated plan.

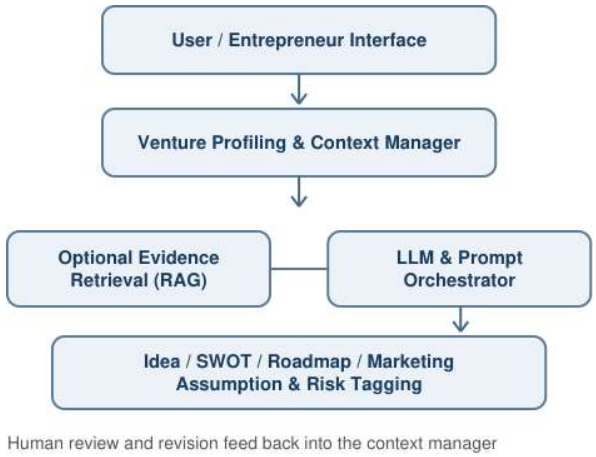


Fig. 1. Proposed modular architecture synthesized from studies [3], [5], [8]–[10].

C. Methodology Flowchart

Figure 2 outlines the proposed workflow: classify the venture stage; elicit and validate context; select an advisory task; retrieve evidence where available; generate structured alternatives; tag risks and uncertainties; and obtain human feedback. Revision may return to context gathering or prompt construction. This is a proposed workflow, not a depiction of a deployed application.

ISSN: 3139-0862



Fig. 2. Proposed venture-stage-aware decision-support flowchart.

D. Prompting, Retrieval and Output Controls

Each prompt would specify the venture objective, known constraints, expected output schema and prohibited unsupported assertions. The optional RAG layer would attach retrieved source titles and dates to verifiable market statements, reflecting [9]. Proposed output fields include recommendation, rationale, evidence status, assumptions, potential risks, validation action, estimated effort and next milestone. The design does not require the graph-based predictive model in [10]; it adopts the more general principle of transparent risk-factor consideration.

E. Human Validation and Lean Iteration

Following [8], recommendations are treated as hypotheses to test through customer interviews, landingpage experiments, prototype feedback or small-scale pilots. A founder can reject suggestions, correct assumptions and compare revisions. This step is especially important for financially consequential claims, legal matters and uncertain market conditions. Public sentiment signals inspired by [6], [7] may be presented as supplementary evidence, never as definitive proof of demand.

F. Proposed Advisory Outputs and Traceability

Each advisory task should produce a compact set of structured fields: task objective, venture stage, relevant input constraints, alternatives, supporting reasons, evidence references, uncertainty notes, validation actions and a next review date. Idea evaluation may include a preliminary scorecard and SWOT observations as in [3], but scores should be accompanied by textual justification. Blueprintstyle planning inspired by [9] may sequence customer discovery, value proposition, early product design and marketing hypotheses. Growth guidance informed by [1], [2] and [4] could focus on service quality, operating efficiency and organizational readiness rather than merely inventing new startup ideas. Traceability matters because an entrepreneur must be able to distinguish an action suggested due to a stated budget from one inferred by the language model. A revision log should record the changed inputs and resulting recommendations, supporting the iterative logic described in [8].

G. Proposed Retrieval and Risk Controls

The optional evidence layer should retrieve only material relevant to the user's selected business question. A source record would contain title, origin, retrieval date, geographic applicability and a short description of the claim it supports. When retrieval is unavailable, the interface should say so rather than inventing citations. The distinction between retrieved data and generated hypotheses reflects the limits of RAG discussed in [9]. Risks can be grouped into market, resource, execution, competitive, compliance and information-quality categories. These are qualitative prompts for review and do not reproduce the specialized graph or federated model of [10]. Likewise, social-media analyses inspired by [6] and [7] should be represented as directional audience signals and should never substitute for representative customer interviews or actual transaction data.

VI. PROPOSED EVALUATION PLAN

No implementation results are available at this stage. A future evaluation could compare the structured workflow with a generic LLM prompt using the same model and matched business scenarios. Test cases should include new ventures and operating businesses from different industries and budgets. Independent reviewers would score relevance, completeness, feasibility, actionability and clarity on a fivepoint scale, while separately counting unsupported factual claims and tracking time to revise a plan.

For each scenario, reviewers should receive anonymized outputs in randomized order. Report sample size, score distributions, inter-rater agreement, factual-grounding errors and qualitative feedback. An optional ablation could compare responses with and without evidence retrieval. A

short-term rating of business advice must not be represented as a demonstrated increase in startup survival or commercial success.

A. Evaluation Questions and Experimental Design

A rigorous future test should evaluate whether a stageaware, structured process improves the usefulness of advice under identical information conditions. Two balanced scenario groups are proposed. New-startup scenarios would vary customer segment, founder skills, capital and initial assumptions; existing-business scenarios would vary operational constraints, available business data and growth objectives. For each case, a baseline prompt and the structured workflow should use the same underlying model, temperature where controllable, and available facts. Outputs would be anonymized and presented in random order to independent reviewers, reducing preference arising from presentation order. A prewritten scoring guide should distinguish evidence quality from fluency: an attractive plan containing unsupported market numbers should receive a factual-grounding penalty regardless of writing quality. This protocol translates the caution about unverified AIgenerated assessments in [3] and [9] into measurable review procedures.

B. Measures and Reviewer Agreement

Relevance measures how well recommendations address the scenario; completeness concerns whether required sections are present; feasibility asks whether actions are plausible within stated resources; actionability assesses the specificity of next steps; and clarity concerns organization. Each dimension can be scored from one to five, accompanied by a rationale. Separately record the count of unsupported factual claims, the share of claims backed by an identifiable source,

ISSN: 3139-0862

the time needed to make a plan acceptable and the number of revision cycles. Multiple raters should score independently, and their agreement should be reported rather than hidden within a mean. Aggregated quality scores must not be confused with realworld startup outcomes. The framework might help someone write a clearer plan without improving investment returns or venture survival; those are distinct hypotheses requiring longitudinal evidence.

C. Ablation, Sensitivity and Failure Cases

A proposed ablation would compare a generic prompt, structured prompting without retrieval, and structured prompting with retrieval. It could identify whether additional evidence improves factual grounding or merely increases response length. A sensitivity test would remove one profile field at a time to see whether recommendations respond appropriately to changes in budget, business stage or target audience. Failure scenarios should include outdated competitor data, contradictory user constraints, ambiguous sentiment signals and missing industry information. These cases are motivated by the extraction problems in [5], uncertainty in social responses [6], [7], and corpus dependence in [9]. The tests are future work; no such experiment has been carried out for this survey.

D. Interpretation Boundaries

The methods and metrics must remain aligned with the scope of the project. Reported predictive accuracy in Pan et al. [10] concerns their own dataset and architecture and cannot be transferred to a general planning assistant.

Similarly, useful output from [3] or [9] does not independently establish financial accuracy. A credible results section should list tested scenarios, review criteria, evaluation limitations and observed errors before drawing conclusions. For a first implementation, qualitative review and error analysis may provide more trustworthy feedback than claiming a percentage increase in startup success. These boundaries reflect the difference between synthesizing research, designing a system and validating the performance of a deployed product.

VII. DISCUSSION, LIMITATIONS AND ETHICS

The literature indicates that AI can structure entrepreneurial information but faces recurring risks: incomplete context [3], [5], nonrepresentative social signals [6], [7], unreliable external data [9], limited generalizability of predictive models [10], and governance or privacy challenges [1], [2]. The proposed design may also produce generic recommendations, fabricated competitor claims or misleading cost estimates. Retrieval itself may surface outdated or biased material. Sensitive financial, customer and IP data should be optional, access-controlled and retained only with informed consent.

A further limitation of this review is its purposive set of ten references rather than a reproducible exhaustive search. The architecture is a synthesis of reported approaches; it is not a replication of the systems in [3], [9] or [10]. In particular, no knowledge-graph classifier, federated-learning pipeline or measured startup-success predictor is claimed for the proposed system.

A. Privacy, Bias and Responsible Deployment

The envisioned application may solicit budgets, operating constraints, customer information and intellectual-property ideas. Collection should be limited to data required for a specific task; access and retention must be controlled. Sensitive records should not be added automatically to external retrieval databases or included in logs without disclosure. Bias may emerge from skewed online information, oversampled customer communities and model assumptions about what a successful business looks like. The governance concerns raised in [1] and [2] therefore apply to individual founders as well as organizations. When sources disagree, outputs should identify the conflict. Where advice crosses legal, medical, tax or investment boundaries, a responsible system should direct users to qualified professionals rather than imply definitive expertise.

B. Transferability and Generalization

The same startup question may have different answers across cities, countries, sectors and stages of maturity. Text sentiment from [6] and [7] is particularly sensitive to platform participation and language context. Practices from educational software ventures [8] may require adaptation to manufacturing, retail or regulated services. The limited sample in [10] shows why even an accurate experimental classifier may not generalize to different venture populations. Further research should therefore diversify scenarios and report subgroup performance, rather than treating a single category of business as representative. The conceptual architecture should preserve enough modularity to add domain-specific validation rules and revise prompts without rewriting the entire application.

C. Practical Implementation Sequence

A defensible next-semester implementation would begin with a venture-stage selector and structured profile form. The second milestone would add a prompt orchestration service that generates consistent advisory sections from a known schema. A third milestone could store user edits and revision history while minimizing retention of sensitive information. Only after these basic functions are tested should a source-linked evidence module be introduced. Acceptance tests would verify that missing fields trigger clarification, unsupported claims remain labeled, citations point to real retrieved records, and a changed user constraint changes the plan appropriately. Evaluation would then use the blinded scenario protocol outlined above. This sequence is a proposal inferred from [3], [5], [8] and [9]; it does not claim the system has already been built.

D. Provenance, Claim Typing and User Trust

A practical output policy can distinguish four claim types: direct user statements, verified quotations or measurements, synthesized interpretations based on cited material, and speculative proposals for future validation. The distinction matters across the reviewed literature. The structured evaluator in [3] produces assessment material that may appear objective because of its scorecard; a clearly labeled inference should not be confused with observed demand. NLP extraction in [5] can help transform a user's narrative into structured attributes but cannot guarantee that the extracted attributes are correct. The retrieval pipeline in [9] can make a generated blueprint more grounded, but it remains necessary to inspect whether referenced material actually entails the presented assertion. The predictive approach in [10] illustrates another possible confusion: a probability generated by a model is

ISSN: 3139-0862

conditional on its training context, not a universal fact about an enterprise. The proposed framework would attach claim-type labels to important assertions and provide a simple method for correcting inaccurate information. These design requirements emerge from analytical synthesis and remain untested.

E. Illustrative Failure and Recovery Scenarios

Consider a founder entering a low budget but requesting a costly market-entry campaign. A generic assistant may comply by inventing a spending forecast. Under the proposed workflow, the system would flag the conflict, explain the assumption and ask whether to adjust the budget or choose a smaller experiment. For an existing business, a claim that customer dissatisfaction is increasing might arise from a narrow sample of social-media posts. Insights inspired by SIRA [6] or fuzzy-sentiment classification [7] would be reported as tentative signals until the firm checks actual customer records. If a retrieved competitor article is old, the source-aware pipeline motivated by [9] would identify its date and qualify the recommendation. If a risk assessment appears numerically precise, the generalization limitations discussed in [10] would justify a warning about sample and population differences. These examples are hypothetical tests of the proposed design; they do not describe observed experimental cases.

F. Assumptions in the Research Synthesis

The ten-paper corpus is intentionally heterogeneous. A study about organizational AI adoption may offer

governance principles but not evidence about a founder's pitch. A language-processing study may explain how to structure an input but not whether subsequent advice yields profitable operations. A startup-oriented retrieval study may produce coherent plans while leaving open questions about the accuracy of market conditions. Therefore the bridge from source contribution to system feature is explicitly inferential. The paper does not assume shared datasets, common evaluation scales or direct reproducibility across the studies. Its contributions are the taxonomy of capabilities, the identification of gaps and the formulation of a testable design. Later development should keep a traceability table mapping each requirement to its source motivation, implementation status and eventual validation evidence, so conceptual inspiration is not misreported as empirical confirmation.

G. Future Research Opportunities

Future experiments can examine whether venture-stageaware interaction improves clarity more for prospective founders than for experienced business owners. Another study could compare revisions made with and without explicit assumption labels, measuring how often users detect unsupported claims. Evidence retrieval can be studied using controlled sets of current and outdated documents, allowing researchers to separate retrieval coverage from generation quality. Cross-sector tests can probe whether outputs designed around technology startups remain suitable for service firms and small local businesses. An optional qualitative interview study could explore whether founders understand uncertainty labels and whether they perceive generated scores as overly authoritative. Longitudinal tracking would be necessary before drawing conclusions about venture performance. The reviewed sources [1]-[10] motivate these directions but do not collectively answer them; each should be framed as a future research question.

H. Risk Categories and Mitigation Matrix

A design-level risk register can distinguish wrong input extraction, missing venture context, fabricated market assertions, stale retrieval, sensitive-data exposure, and overconfident predictions. Input extraction errors may be mitigated through editable profiles and explicit confirmation, following the architectural motivation of [5]. Missing context can prompt clarification rather than guesswork. Fabricated assertions require citations or visible uncertainty labels, a concern directly relevant to structured generation [3] and retrieval-augmented blueprints [9]. Stale market evidence should carry a retrieval timestamp and region label. Sensitive business information requires minimization, access controls and disclosure of processing boundaries in line with adoption and governance considerations [1], [2], [4]. Prediction overconfidence should trigger model-scope warnings rather than presenting risk-classification research [10] as a deployed startupsuccess oracle. The matrix is intended as a proposed engineering artifact, not as an empirically measured reduction in error.

I. Structured Comparison of Entrepreneurial Outputs

The reviewed approaches imply several distinct categories of useful output. A venture idea scorecard inspired by [3] should identify the assumptions behind every rating. A venture profile inspired by [5] should summarize customers, constraints and resources without inventing

missing details. Market-signal summaries motivated by [6] and [7] should explain collection bias and sentiment ambiguity. Lean tasks inspired by [8] should describe a falsifiable question, a practical test and a decision threshold. A blueprint informed by [9] should connect recommendations to dated sources whenever possible. Risk notes motivated by [10] should name potential failure mechanisms and disclose that qualitative assessments are not validated success probabilities. Combining the categories in one interface may help users compare alternative plans, yet a future evaluation must determine whether the extra structure aids decision-making or merely increases reading burden.

J. Functional and Non-Functional Requirements

The literature can also be translated into testable requirements. A functional requirement is the ability to classify a request as ideation or growth and present the corresponding questions; another is to produce multiple alternatives with editable assumptions. Structured scoring and decision artifacts are motivated by [3], while consistent venture profiling draws from the information-extraction perspective of [5]. Retrieval may be optional, but when enabled it should display source metadata and separate an evidence excerpt from a model-generated recommendation [9]. A feedback function should preserve changes between iterations, consistent with lean experimental practice [8]. Non-functional requirements include privacy, usability, transparent error handling, reproducibility of scenario tests and resilience when a retrieval service is unavailable. These expectations reflect governance concerns in [1] and [2]. None is presented as a completed implementation or a measured performance property; each can be checked through an acceptance test in the following semester.

ISSN: 3139-0862

K. Example Scenario: A First-Time Founder

An illustrative participant might propose a low-cost service aimed at students but lack evidence that a target group will pay. A stage-aware profile would collect available skills, geographic reach, approximate budget, target customers and the user's initial problem statement. Inspired by the structure in [3], the advisory response could compare several value propositions and list qualitative strengths and weaknesses for each. Using the lean methodology motivation in [8], the response would recommend a small number of interviews and an inexpensive prototype before substantial spending. A retrieval-supported extension inspired by [9] might identify dated external information but should explicitly mark uncertain prices or competitor details. The entrepreneur could then return with interview findings and revise the roadmap. The scenario illustrates how the conceptual components fit together; it is not an empirical case study and no customer demand, revenue or conversion outcomes are assumed.

L. Example Scenario: An Operating Business

A different illustrative user might manage a small established business experiencing weaker repeat purchases. The interface would ask about existing products, channels, customer feedback, staffing constraints and the definition of the business problem. AI-adoption and governance literature [1], [2] motivates attention to organizational readiness, while the growth and IP-strategy perspective [4] motivates considering differentiation and knowledge assets. Public-

review signals could be summarized cautiously using concepts from [6] and [7], yet real transaction records and direct customer feedback should take precedence when available. Possible outputs might include service-process hypotheses, low-cost retention experiments and a phased action plan. A source and uncertainty label would prevent the tool from converting unsupported speculation into a factual diagnosis. The exercise underscores that existingbusiness improvement differs from unrestricted ideation even when both rely on similar language-model infrastructure.

M. Limitations of Performance Comparison

Because the surveyed studies target different problem formulations, they cannot be pooled into one accuracy estimate. Sentiment-classification performance on socialmedia text [7] would describe a different output and ground truth than a knowledge-graph success classifier [10]. Narrative recommendations from an idea evaluator [3] or startup blueprint generator [9] may be rated by humans, but such ratings are not directly comparable to model discrimination metrics. Organizational adoption insights in [1] and [2] may have no single numerical score at all. Therefore the review uses qualitative synthesis rather than a meta-analysis of effect sizes. The proposed framework should be evaluated against clearly specified baselines on matched inputs and with disclosed rater instructions. Even strong scores would support only the tested tasks and populations. This distinction helps avoid misleading readers about both the cited studies and the present unimplemented proposal.

N. Reproducibility and Documentation Plan

For future evaluation, each test scenario should have a unique identifier, an input profile, a declared task type and an expected list of output fields. Prompts, model version, retrieval settings and source dates should be recorded to make repeated trials interpretable. Reviewer instructions should define relevance, completeness, feasibility, actionability and clarity before scoring begins. Qualitative disagreements can be logged along with concrete failure examples, and revisions to the system should be accompanied by renewed baseline comparisons. The traceability principle extends to literature citations: a claimed design requirement can cite its motivating study, while its actual effectiveness must be supported by newly collected evidence. Such documentation would help distinguish the methodological proposal in this paper from later system demonstrations and results. The planned procedure also makes it possible to report null or unfavorable findings without concealing them behind polished generated text.

O. Proposed Acceptance Testing

A minimal acceptance-test suite would target observable behavior rather than subjective enthusiasm for AI-generated prose. One test would verify that selecting an operating venture produces questions about current customers, operations and constraints rather than asking only for a novel startup idea. Another would check that missing financial data are represented as missing and do not become invented revenue estimates. A third would verify that switching a founder's budget changes recommended tactics and exposes the reason for the revision. A fourth would inspect source records to ensure that citations identify actual

retrieved documents rather than model-generated titles. A fifth would confirm that a user can delete or correct sensitive profile fields without them resurfacing in later outputs. These tests connect the structured-evaluation idea in [3], the extraction and normalization motivation in [5], the retrieval orientation of [9] and the governance concerns discussed in [1] and [2]. They remain proposed checks, not results obtained from a functioning platform.

P. Human-in-the-Loop Decision Boundaries

Automation should support rather than replace entrepreneurial judgment. A founder is best positioned to confirm local market context, financing restrictions and willingness to run a proposed experiment. The system can organize choices but should not approve a loan, assert legal compliance or predict venture survival from an unverified narrative. Consistent with lean iteration [8], the decision process should end with a test, a measurement method and an explicit opportunity to revise assumptions. If the system presents a market opportunity score inspired by [3], it should indicate which parts are user-supplied and which are speculative. If it includes risk factors motivated by [10], it should present them as considerations for review rather than probabilities inferred from an unimplemented prediction model. This boundary is essential for transparency and for avoiding unwarranted trust in fluent model outputs.

Q. Interpretation of the Ten-Paper Evidence Base

Taken together, the reviewed papers motivate a modular advisory architecture but do not prescribe a single optimal implementation. Studies [1], [2] and [4] emphasize organizational setting and adoption; [3] motivates structured startup assessments; [5] illustrates a method for handling natural-language descriptions; [6] and [7] address noisy public responses; [8] supplies an iterative validation perspective; and

ISSN: 3139-0862

[9] and [10] demonstrate specialized directions in retrieval-supported generation and risk modeling. The proposed project treats those findings as complementary design inputs rather than combining unrelated outcome numbers. Because review selection was purposive, the synthesis may omit competing approaches and newer evidence. Any claim that the integrated design performs better than a general assistant should be tested independently against a documented baseline. By making this limitation explicit, the paper seeks to provide a realistic research starting point for future implementation and evaluation.

R. Scope of Future Empirical Conclusions

A future experiment could establish whether recommendations are more complete, easier to revise or better grounded in identified evidence for the tested tasks. It could also reveal that a simpler baseline performs similarly, or that additional structure increases time and complexity without meaningful benefit. Both outcomes should be reported. Scores from a short laboratory study would not prove that the framework improves startup investment, long-term growth or survival. Such outcomes depend on industry conditions, founder decisions, customer behavior and many factors beyond software-generated advice. Likewise, a small convenience sample cannot demonstrate validity across all geographical regions and business sectors. A cautious interpretation would connect each conclusion to its measurement instrument, participant group and time horizon. This boundary keeps the present paper aligned with

its central identity as a ten-study survey and a proposal for subsequent research.

VIII. CONCLUSION AND FUTURE WORK

This survey reviewed ten studies relevant to AI-supported entrepreneurship and mapped their contributions to idea evaluation, organizational adoption, NLP, sentiment analysis, lean experimentation, blueprint generation and startup risk assessment. It proposed a stage-aware advisory architecture and flowchart with structured outputs, optional evidence retrieval, uncertainty labeling and iterative human review. Future work should implement a minimum viable version, verify the accuracy of sourced recommendations and evaluate usefulness across realistic new-venture and existing-business scenarios. Any performance or impact claims must follow actual empirical testing.

REFERENCES

[1] C. Rios-Campos et al., “Startups and Artificial Intelligence,” South Florida Journal of Development, vol. 5, no. 2, pp. 950–969, 2024, doi: 10.46932/sfjdv5n2042. [2] M. Al-Kfairy, “Strategic Integration of Generative AI in Organizational Settings: Applications, Challenges, and Adoption Requirements,” IEEE Engineering Management Review, vol. 53, no. 6, 2025, doi: 10.1109/EMR.2025.3534034. [3] A. Davalas and A. Angelaki, “Startup Idea Evaluator: Supporting and Evaluating Early-Stage Startup Ideas Using ChatGPT,” International Journal of Social Science and Economic Research, vol. 10, no. 1, 2025, doi: 10.46609/IJSSER.2025.v10i01.024.

[4] B. H. Reddy, K. V. Datta, C. Naveen, D. S. S. Bhaswanth, and M. N. Vijayakumar, “The Role of AI in Empowering Startup Growth and IP Strategy: Case Studies & A Practical Web Platform,” research manuscript supplied for literature review; full publication details not verified. [5] P. Danenas and T. Skersys, “Exploring Natural Language Processing in Model-To-Model Transformations,” IEEE Access, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3219455. [6] B. Alotaibi et al., “Startup Initiative Response Analysis (SIRA) Framework for Analyzing Startup Initiatives on Twitter,” IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2965181. [7] H. T. Phan, V. C. Tran, N. T. Nguyen, and D. Hwang, “Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model,” IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2019.2963702. [8] O. Cico, “Lean Software Startup Practices and Software Engineering Education,” in Proc. IEEE/ACM 44th Int. Conf. Software Engineering: Companion Proceedings, 2022, doi: 10.1109/ICSECompanion55297.2022.9793731. [9] Z. Mansuri, S. Shaikh, B. Shaikh, P. Jagtap, and S. Gurav, “AI-Driven Startup Blueprint Generator using Retrieval-Augmented Generation and Large Language Models,” Multidisciplinary Journal of Research in Engineering and Technology, vol. 13, no. 1S, pp. 17–23, 2026, doi: 10.65521/mjret.v13i1S.3023. [10] C. Pan, X. Pan, and L. Sun, “A Multi-Layer AI Decision Support System for Startup Success Prediction and Risk Assessment Using Knowledge Graphs and Federated Learning,” Scientific Reports, vol. 16, art. no. 20259, 2026, doi: 10.1038/s41598-026-44162-8.