© Volume 2, Issue 10, October 2026 | JATIR
e-ISSN: 2839-5786, p-ISSN: 3139-8464

AI Startup Validation Platform an Intelligent System for Automated Startup Idea Evaluation and Feasibility Analysis

[1]Yash Amar Andhare, [2]Pranav Santoshrao Kharat, [3]Yashwant Bhaskar Mane
[4]Prof. Komal Palikondawar
[1,2,3]Department of Information Technology, Genba Sopanrao Moze Trust's Parvatibai Genba Moze College of Engineering, Wagholi, Pune, India
[4]Professor, Department of Information Technology, Genba Sopanrao Moze Trust's Parvatibai Genba Moze College of Engineering, Wagholi, Pune, India

***Abstract*—Most startups fail early, often because ideas were never validated in a structured, unbiased way. Founders rely on gut feeling, friends and family, or costly consultants. This work proposes a platform that combines NLP, ML classifiers, LLM reasoning, and live external data (market trends, startup and funding databases). It takes a text description of an idea and returns a feasibility report. The report has five sub-scores (market demand, competition, originality, technical feasibility, revenue potential), a composite score, a SWOT analysis, and improvement suggestions. The goal is to cut validation time from weeks to minutes for students, first-time founders, and incubators.**

***Index Terms*—Startup validation, idea evaluation, natural language processing, large language models, machine learning, feasibility analysis, SWOT analysis.**

I. INTRODUCTION

Starting a company has become a popular career choice in India, supported by programmes such as Startup India, incubation cells and a growing venture capital scene, yet most startups still fail within their first few years, and weak or missing idea validation is one of the main causes. First-time founders and student teams usually validate ideas informally, by asking friends, running small surveys or trusting their gut, while professional market research is out of reach for them because of its cost. Advances in NLP, large language models and machine learning now make it possible to read a free-text idea, compare it with existing companies, check market trends and produce a structured assessment in minutes, but earlier projects mostly rely on a fixed dataset, give a single unexplained score, or cover only a few industries. This paper presents an AI Startup Validation Platform for students, first-time founders and incubators, and its main contributions are a hybrid

scoring approach that combines machine learning with LLM reasoning, the use of live market and funding data, an explainable output with five separate sub-scores, a SWOT analysis and practical suggestions, and an evaluation plan that follows the SAISE framework.

II. PROBLEM STATEMENT

First-time entrepreneurs and student innovators typically lack access to professional market research, mentorship, or investor-grade due diligence when forming a new business idea. Existing validation is largely manual — informal surveys, personal networks, or paid consultants — and is slow, subjective, expensive, and inconsistent across ideas and evaluators. There is no easily accessible, automated, and objective platform that can take a startup idea as input and return a consistent, evidence-based assessment of its market potential, competition, feasibility, and originality within a short time frame. This gap results in founders investing time and money in ideas that were never viable, while genuinely promising ideas may be abandoned due to a lack of confidence or guidance.

III. MOTIVATION

- The high startup failure rate (cited as up to 90% within the first year) is disproportionately linked to inadequate idea validation rather than execution failure alone.
- Professional startup consultancy and market research are costly and largely inaccessible to students and early-stage founders.
- The rapid maturity of NLP and LLM technologies now allows automated, near real-time analysis of unstructured business ideas at a scale and speed not possible with manual research.
- Academic incubators and entrepreneurship cells need a lightweight, repeatable tool to triage and mentor a large number of student ideas fairly and efficiently.
- Existing commercial idea-validators are largely proprietary, subscription-based, and not tailored to the Indian startup context, motivating an academically grounded, transparent alternative.

IV. AIM AND OBJECTIVES

Aim: To design and implement an AI-based platform that automatically validates a startup idea and provides founders with a transparent, data-driven feasibility report.

Objectives:

- To study existing approaches to AI-driven startup evaluation and idea validation through a structured literature survey.
- To design a system architecture that accepts a natural-language startup idea description as input and processes it through NLP and machine learning pipelines.

- To integrate external data sources (e.g., market trend and startup/funding data) to ground the analysis in real-world evidence rather than the model's own assumptions alone.
- To generate a composite feasibility score along with a SWOT analysis and specific, actionable improvement suggestions for each submitted idea.
- To build a usable web interface so that students, entrepreneurs, and incubators can submit ideas and view/download validation reports.
- To evaluate the system's outputs for consistency and usefulness against a sample set of real and hypothetical startup ideas.

V. SCOPE OF THE PROJECT

- Works from a plain-text description of a startup idea supplied by the user, and produces an early-stage, pre-launch feasibility assessment — it is not meant to replace full due diligence carried out by investors or professional consultants.
- Evaluates market demand, competitive landscape, originality, technical feasibility, and indicative revenue potential, drawing on a mix of publicly available and simulated datasets.
- Supports any industry or domain the user describes, though the depth of the evaluation depends on how much data is available for that particular domain.
- Is delivered as a web-based application, with the ability to generate and export a report.
- Leaves financial forecasting, legal/regulatory compliance checking, and investor matchmaking out of the current version; these are listed separately under Future Scope.

VI. EXISTING SYSTEM AND ITS LIMITATIONS

Current idea-validation approaches can be grouped into three broad categories:

1. Manual/consultative validation — founders rely on personal networks, mentors, or paid consultants. This route is slow, expensive, and subjective, with no standard scoring method.
2. Generic AI idea-validator products — consumer web tools that generate market and feasibility reports using LLMs. These tend to be closed-source, subscription-gated, and not adapted to local or academic contexts.
3. Academic prototypes reported in the literature — for example, Kumar et al. (IJEAT, 2020) trained Linear Regression and Support Vector Machine models on an expert-curated dataset of over 50 parameters, while more recent Indian student projects such as “AI Startup Idea Validator Using ML and LLM Agents” and “IntelliStart” combine LLM-based reasoning with market and trend data to produce feasibility scores and SWOT reports.

Common limitations identified across these systems include:

- Heavy reliance on a single static dataset, which can become outdated and does not reflect current market trends.
- Limited or no integration of live external signals such as search-trend or funding-database data.

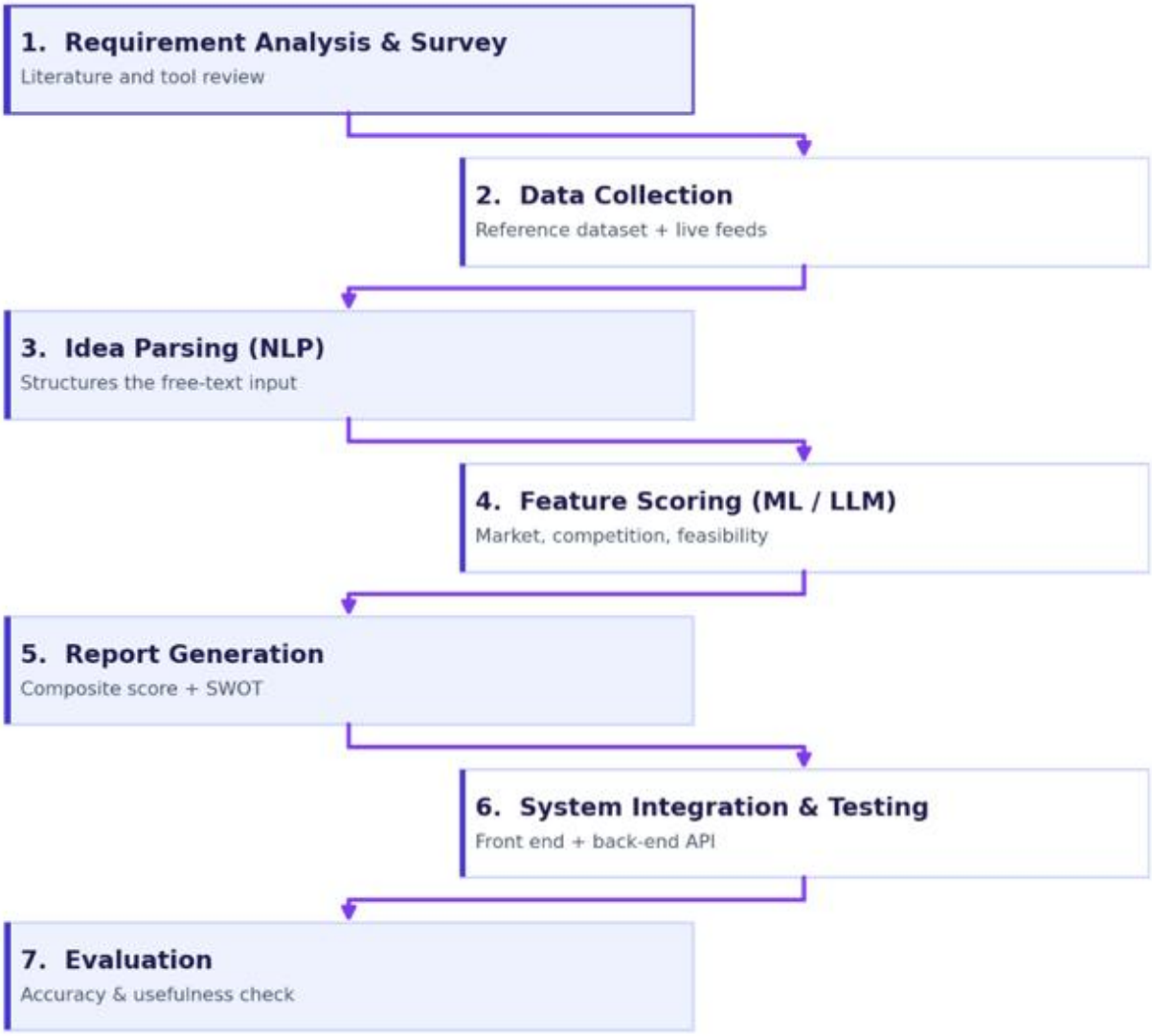
- Feasibility scores that are often opaque, with little explanation of how a score was derived (low interpretability).
- Narrow domain coverage — many prototypes are tuned to a small set of industries and generalize poorly.
- No standardized evaluation methodology across studies, which the SAISE framework (Jafari et al., 2025) identifies as a major gap limiting comparability and reliability of AI-driven startup evaluation research.

VII. PROPOSED SYSTEM

The proposed AI Startup Validation Platform addresses these limitations by combining an NLP-based idea parser, a multi-factor ML/LLM scoring engine, and live external data integration (market trend and startup/funding signals) within a single, transparent web platform. Unlike the reviewed prototypes, the system is designed to explain each component of its composite score (market demand, competition, originality, feasibility, and revenue potential) individually, and to generate a human-readable SWOT analysis and improvement suggestions rather than a single opaque number. The platform targets students, first-time entrepreneurs, and academic incubators as primary users, and is built to be extensible to new domains and data sources as they become available.

VIII. PROPOSED METHODOLOGY

- Requirement analysis and literature survey of existing AI-driven startup evaluation systems and datasets.
- Data collection: assembling a training/reference dataset of startup descriptions, outcomes, and relevant market indicators, supplemented with real-time data pulled from external sources (e.g., trend and funding/startup databases).
- Idea parsing: using NLP techniques (tokenization, embeddings, and/or LLM-based summarization) to convert the user's free-text idea description into structured features (domain, target customer, problem, proposed solution).
- Feature scoring: applying machine learning classifiers/regressors (e.g., Logistic Regression, SVM, ensemble models) and/or LLM-based reasoning to score market demand, competition, originality, feasibility, and revenue potential.
- Report generation: aggregating individual scores into a composite validation score, generating a SWOT analysis, and producing improvement suggestions.
- System integration and testing: connecting the NLP/ML pipeline to a web front end and back-end API, followed by functional, usability, and accuracy testing against sample startup ideas.
- Evaluation: comparing platform outputs against known outcomes of real startups (where available) and expert/peer review to assess consistency and usefulness.

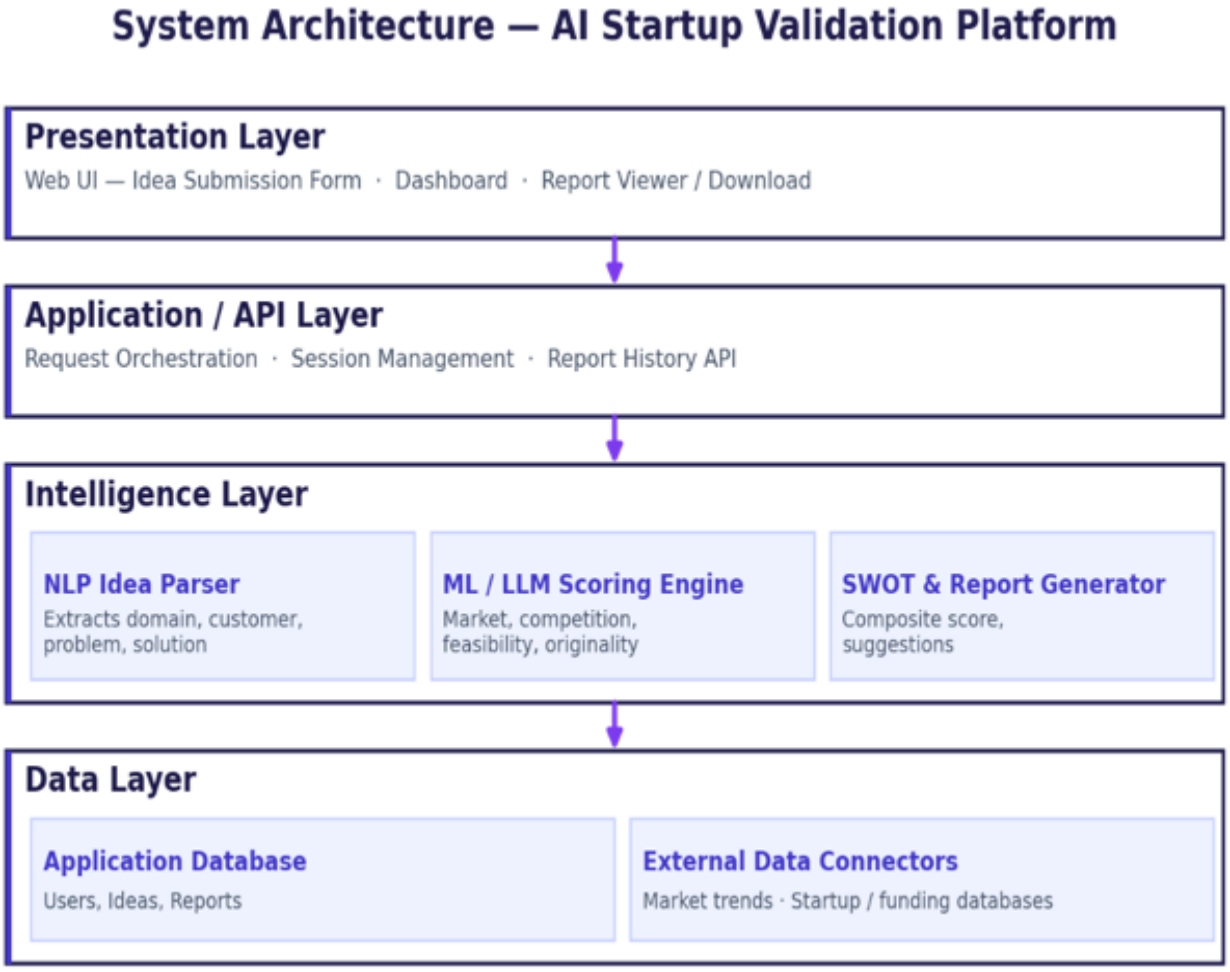


IX. SYSTEM ARCHITECTURE

The system follows a layered architecture made up of four principal layers:

1. Presentation Layer — a web-based front end where users submit their startup idea description and view or download the generated report.
2. Application/API Layer — receives the request, orchestrates calls to the NLP and scoring modules, and manages user sessions and report history.
3. Intelligence Layer — contains the core NLP parser, the ML/LLM-based scoring engine, and the SWOT/suggestion generator, and is responsible for producing the structured evaluation.
4. Data Layer — comprises the application database (user accounts, submitted ideas, generated reports) together with connectors to external data sources such as market-trend feeds and startup/funding databases, which supply the real-world signals used during scoring.

Data flows from the presentation layer through the API layer into the intelligence layer, which queries the data layer for both stored reference data and live external data before returning the finished report back through the API to the presentation layer.



X. MODULE DESCRIPTION

A. User Interface Module

Provides idea-submission forms, dashboards for viewing past evaluations, and report export/download functionality.

B. Idea Parsing (NLP) Module

Extracts structured attributes — domain/industry, target customer segment, core problem, and proposed solution — from the user's free-text description using NLP and/or LLM-based summarization.

C. Market & Competitor Analysis Module

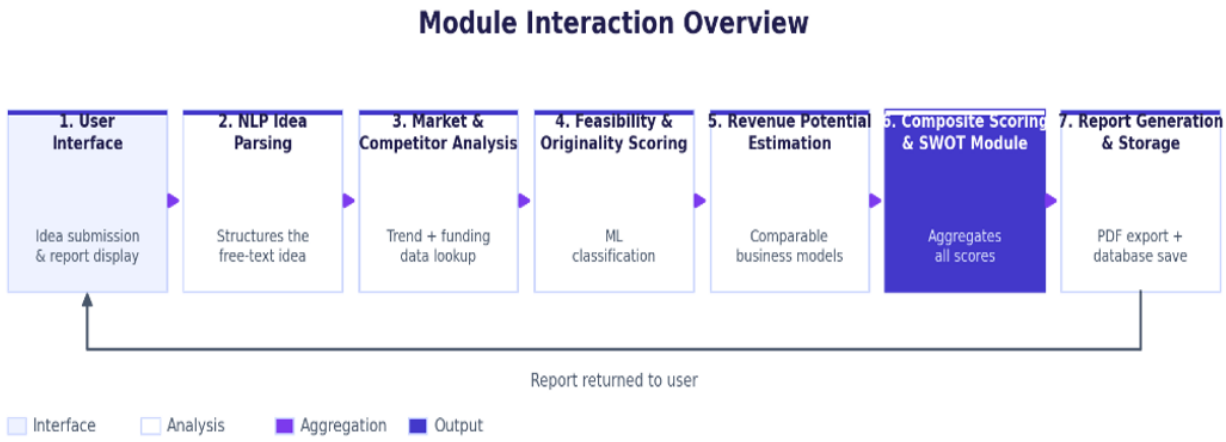
Queries external trend and startup/funding data sources to estimate market size, growth trend, and the density/strength of existing competitors in the same domain.

D. Feasibility & Originality Scoring Module
Applies machine learning models to rate technical feasibility and the originality/novelty of the idea relative to the reference dataset.

E. Revenue Potential Estimation Module
Provides an indicative estimate of monetization potential based on comparable business models and market size.

F. Composite Scoring & SWOT Module
Aggregates individual module outputs into a single composite validation score and produces a structured SWOT (Strengths, Weaknesses, Opportunities, Threats) summary with improvement suggestions.

G. Report Generation & Storage Module
Compiles all outputs into a downloadable report (PDF/web view) and stores it against the user's account for future reference.



XI. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware Requirements (Development):
- ❖ Processor: Intel i5 (or equivalent) or higher
- ❖ RAM: 8 GB minimum (16 GB recommended for local model experimentation)
- ❖ Storage: 20 GB free disk space
- ❖ Internet connectivity for access to external data sources and cloud/LLM APIs

Software Requirements:
- ❖ Operating System: Windows 10/11, macOS, or Linux

- Programming Languages: Python (NLP/ML pipeline), JavaScript/TypeScript (web front end)
- Frameworks/Libraries: scikit-learn, pandas, NumPy, an NLP/LLM library or API (e.g., spaCy, Hugging Face Transformers, or an LLM API), React/Node.js for the web application
- Database: MySQL / PostgreSQL / MongoDB
- Version Control & Deployment: Git/GitHub, Docker (optional), a cloud platform (e.g., AWS/Azure/Render) for hosting

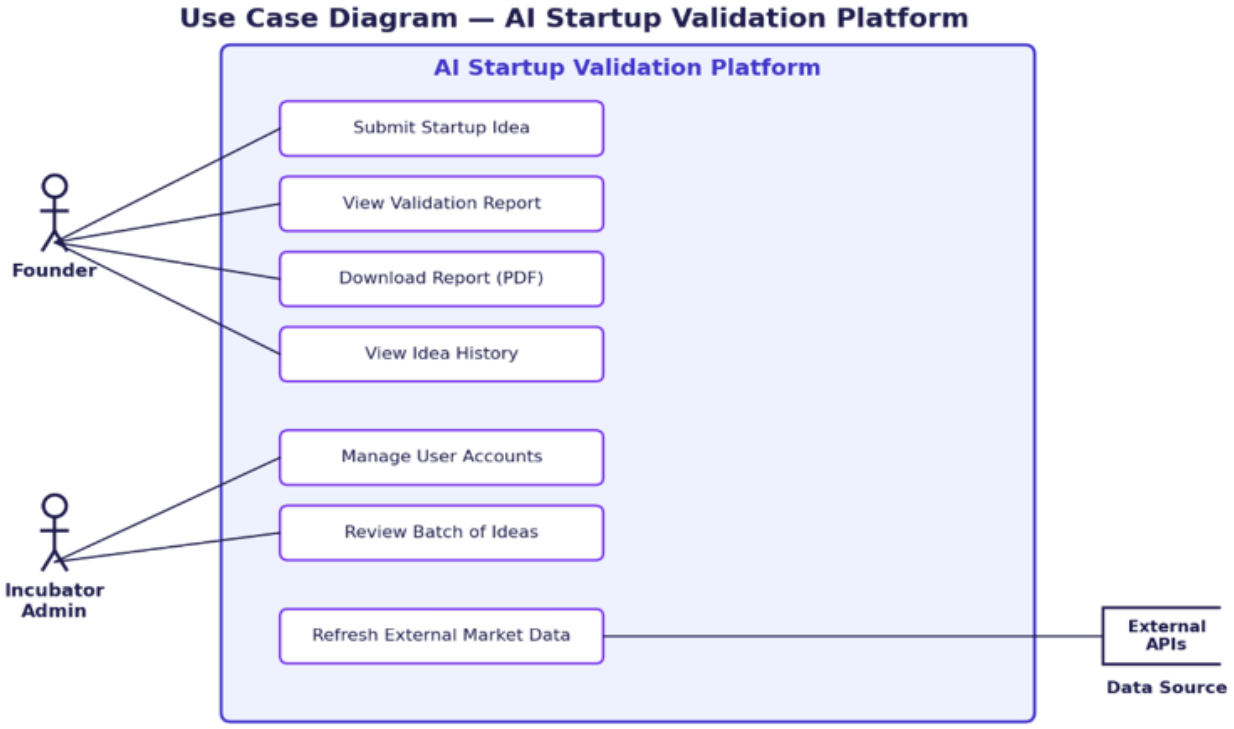
XII. DATABASE/UML DESIGN, WHEREVER APPLICABLE

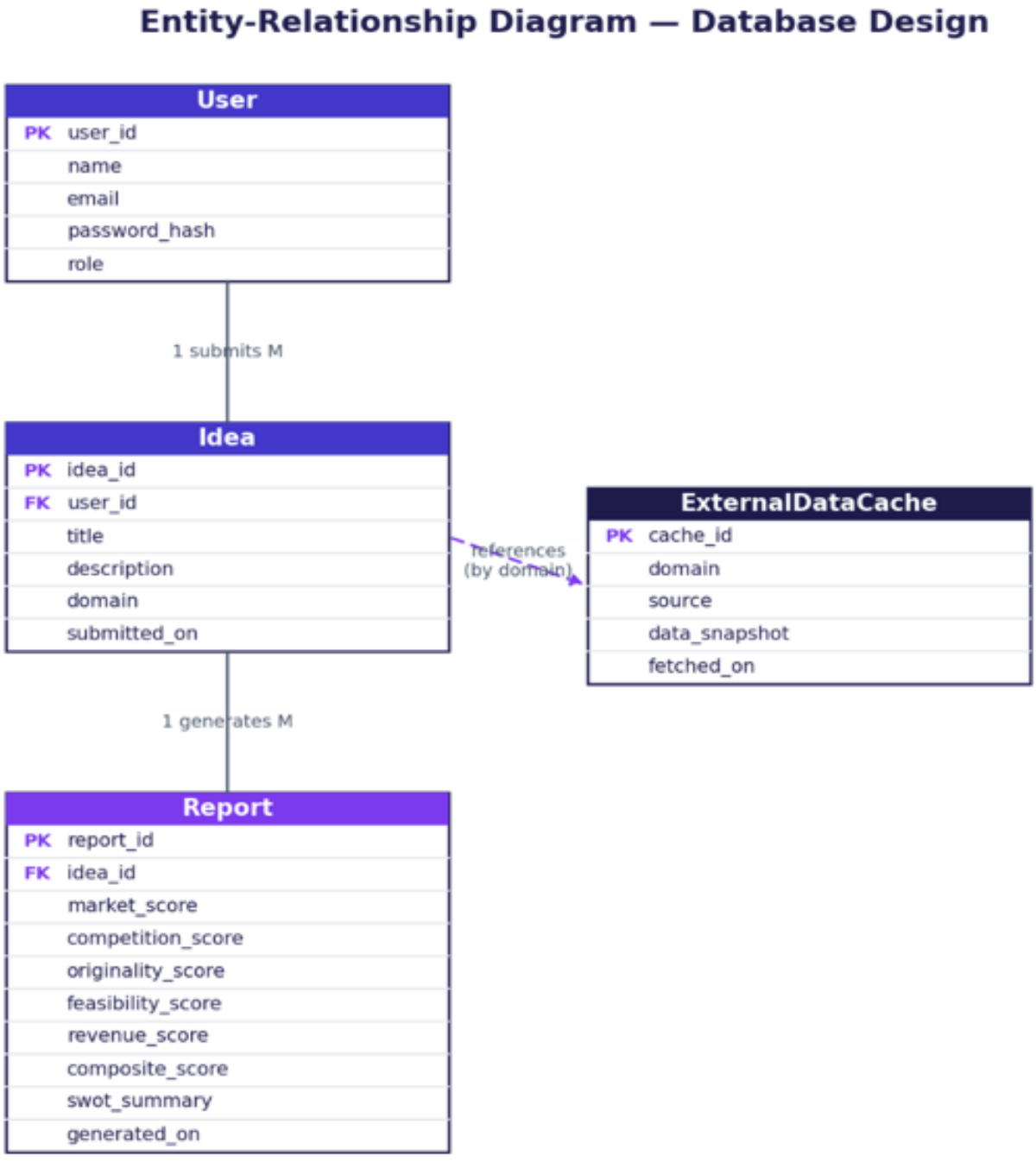
The relational schema will include, at minimum, the following entities:

- User (user_id, name, email, password_hash, role)
- Idea (idea_id, user_id, title, description, domain, submitted_on)
- Report (report_id, idea_id, market_score, competition_score, originality_score, feasibility_score, revenue_score, composite_score, swot_summary, generated_on)
- ExternalDataCache (cache_id, domain, source, data_snapshot, fetched_on), used to store periodically refreshed external market/competitor data

UML diagrams prepared during the design phase include:

- A Use Case diagram covering the Founder, Incubator Admin, and Data Source actors.
- An Entity-Relationship diagram reflecting the schema above.
- A Sequence diagram illustrating the flow from idea submission to report generation (to be elaborated during detailed design).
- An Activity diagram for the internal scoring pipeline (to be elaborated during detailed design).





XIII. INNOVATION/NOVELTY

- Combines traditional supervised ML scoring with LLM-based reasoning, rather than relying on either approach alone, to balance interpretability with reasoning depth.
- Integrates live external market and competitor signals, rather than relying solely on a static training dataset, keeping evaluations current.

- Produces an explainable, component-wise feasibility score with an accompanying SWOT analysis and concrete improvement suggestions, rather than a single opaque number.
- Purpose-built for students, first-time founders, and academic incubators, an audience underserved by existing subscription-based commercial validators.

XIV. EXPECTED OUTCOMES

- A working prototype of the AI Startup Validation Platform capable of accepting a startup idea and returning a structured feasibility report within a short response time.
- A composite validation score along with individual sub-scores for market demand, competition, originality, feasibility, and revenue potential.
- An automatically generated SWOT analysis and a set of actionable improvement suggestions for each submitted idea.
- A usable web interface suitable for demonstration to students, mentors, and incubator staff.
- A documented evaluation of the system's scoring consistency against a curated set of test ideas.

XV. APPLICATIONS

- Assisting student entrepreneurs and college innovation/incubation cells in screening and prioritizing project or startup ideas.
- Supporting early-stage founders in refining their idea before approaching investors or accelerators.
- Providing a preliminary screening tool for hackathons and business-plan competitions.
- Assisting mentors and faculty guides in providing consistent, data-backed feedback across a large number of student projects.

XVI. ADVANTAGES AND LIMITATIONS

Advantages:

- Fast, low-cost, and consistent evaluation compared with manual or consultancy-based validation.
- Objective, data-driven scoring that reduces founder bias and overconfidence.
- Accessible to students and early-stage founders who cannot afford professional market research.
- Scalable — capable of evaluating a large volume of ideas simultaneously, useful for incubators and competitions.

Limitations:

- Output quality depends on the availability and freshness of external market/competitor data for a given domain.

- The platform provides an indicative, preliminary assessment and is not a substitute for full investor due diligence or legal/financial review.
- LLM-based reasoning components may occasionally produce inconsistent or overly generic feedback, requiring human review.
- Currently limited to text-based idea descriptions; does not yet process financial models, pitch decks, or prototypes directly.

XVII. FUTURE SCOPE

- Incorporating investor-matching functionality to connect validated ideas with relevant angel investors or venture capital firms.
- Adding automated financial forecasting (TAM/SAM/SOM estimation, break-even analysis) as demonstrated in commercial AI co-founder tools.
- Extending the platform to accept and analyze pitch decks, prototypes, or landing-page traffic data alongside text descriptions.
- Building a feedback loop where actual startup outcomes are fed back into the model to continuously improve scoring accuracy, addressing the reproducibility gaps identified in current AI-driven startup evaluation research.
- Adopting a more principled, standardized evaluation methodology, in line with frameworks such as SAISE, to improve comparability and reliability of results across domains.

XVIII. LITERATURE SURVEY

The literature reviewed for this proposal spans academic studies on machine-learning-based startup evaluation, applied AI validator systems, and recent research on standardized evaluation methodology, together supporting the need for an accessible, data-driven validation tool aimed at students and early-stage founders.

Source	Key Finding / Relevance to the Platform
Kumar et al. [1] — "Startup Idea Validation using Machine Learning" (IJEAT, 2020)	Used Linear Regression and SVM models trained on 50+ expert-curated parameters, reporting high training accuracy; highlighted the absence, at the time, of any comparable automated validation tool — motivating the need for a platform of this kind.
ACE Engineering College, Hyderabad [2] — "AI Startup Idea Validator Using ML and LLM Agents"	Combined ML with LLM agents in a web interface producing a feasibility score, SWOT analysis, and improvement suggestions — closely aligned with this platform's own report-generation approach.

Prajapati & Albert [3] — "IntelliStart: AI-Driven Startup Idea Validation and Market Feasibility System"	Used NLP, trend analysis, and competitor mapping (Google Trends, Crunchbase) aimed at students and early-stage founders lacking access to professional consultancy — supports the target-audience framing adopted here.
Ganuthula & Kuruva [4], IIT Jodhpur / IIT Bombay	A natural-experiment study of 914 AI-era firms found larger structures and higher absolute valuations but lower per-employee efficiency, showing that AI adoption changes firm dynamics a validation platform should account for.
Jafari et al. [5] — SAISE Framework	Proposed a standardized, principled evaluation methodology, identifying its absence as a major gap limiting the comparability and reliability of AI-driven startup evaluation research.

Research gap addressed by the project

For this academic prototype, the practical gap is not the absence of AI-assisted startup evaluation tools in general. Instead, the project focuses on combining ML-based scoring, LLM-based reasoning, and external market/competitor signals into a single, transparent report-generation pipeline aimed specifically at students and early-stage founders, while adopting a standardized evaluation methodology in line with the SAISE framework to address the reproducibility and comparability gaps identified in prior research.

The reviewed studies should be treated as background references for scoring methodology and evaluation design; the platform's own composite score and SWOT output remain an indicative, preliminary assessment and are not a substitute for full investor due diligence or professional review.

REFERENCES

[1]. S. Kumar, B. Guirish, A. Umale, and P. Ganjewar, "Startup Idea Validation using Machine Learning," International Journal of Engineering and Advanced Technology (IJEAT), vol. 9, no. 5, June 2020. Available: https://www.ijeat.org/wp-content/uploads/papers/v9i5/E9549069520.pdf

[2]. "AI Startup Idea Validator Using ML and LLM Agents," Dept. of CSE (AI & ML), ACE Engineering College, Hyderabad, India, International Journal of Scientific Research in Engineering and Technology (IJSRET), vol. 12, issue 2. Available: https://ijsret.com/wp-content/uploads/1775681638_1775651120_IJSRET_V12_issue2_316.pdf

[3]. A. R. C. Prajapati and D. W. Albert, "IntelliStart: AI-Driven Startup Idea Validation and Market Feasibility System," International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), vol. 13, issue 10, 2025, DOI: 10.15680/IJIRCCE.2025.1310014. Available:

https://ijircce.com/admin/main/storage/app/pdf/14_IntelliStart%20AI-Driven%20Startup%20Idea%20Validation%20and%20Market%20Feasibility.pdf

[4]. V. R. R. Ganuthula and R. Kuruva, "AI's Structural Impact on India's Knowledge Intensive Startup Ecosystem: A Natural Experiment in Firm Efficiency and Design," arXiv:2507.19775, July 2025. Available: https://arxiv.org/pdf/2507.19775

[5]. S. M. A. Jafari et al., "Deconstructing the Crystal Ball: From Ad-Hoc Prediction to Principled Startup Evaluation with the SAISE Framework," arXiv:2508.05491, August 2025. Available: https://arxiv.org/pdf/2508.05491