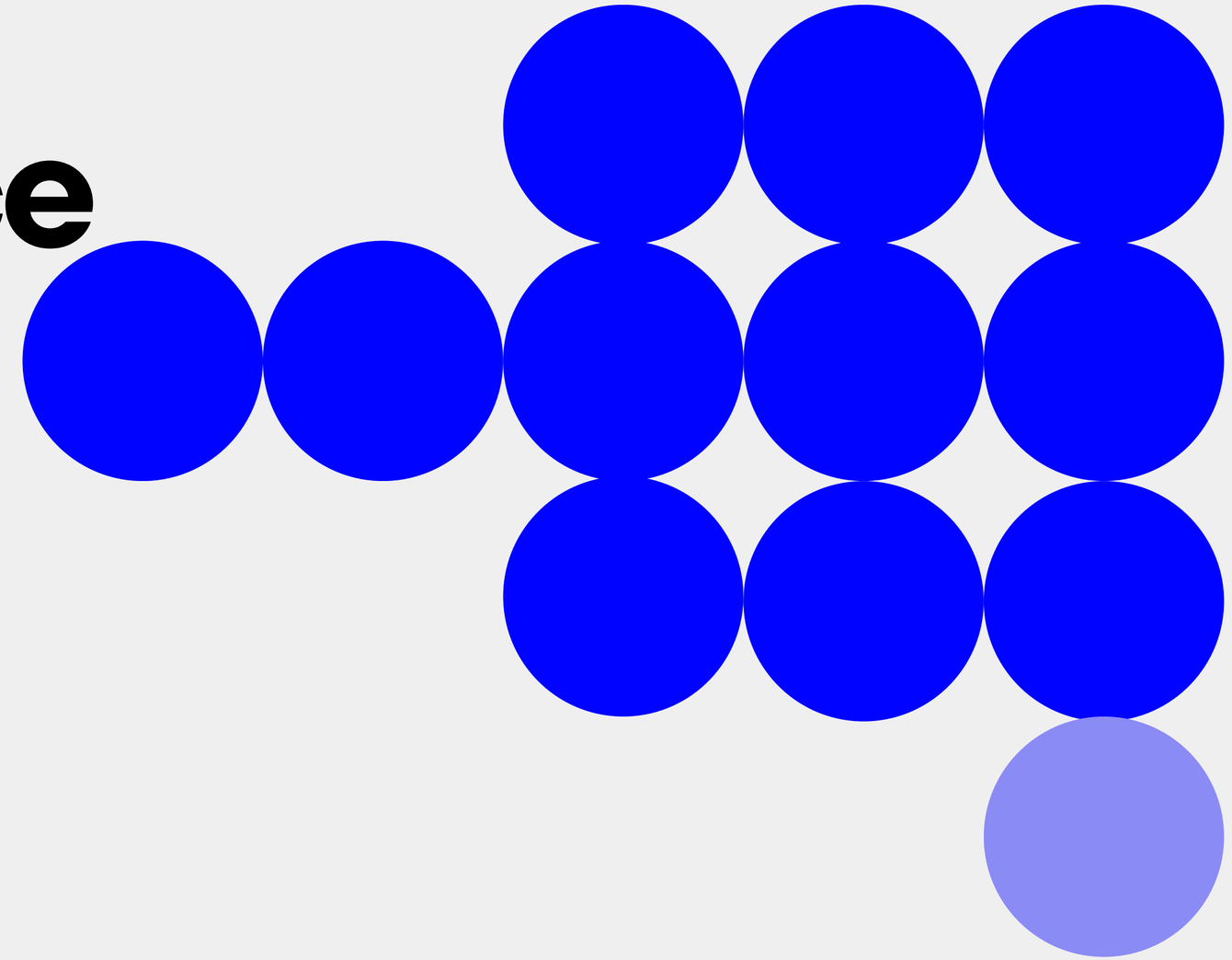
AI-Powered Startup Application Screening

Quintessence AI

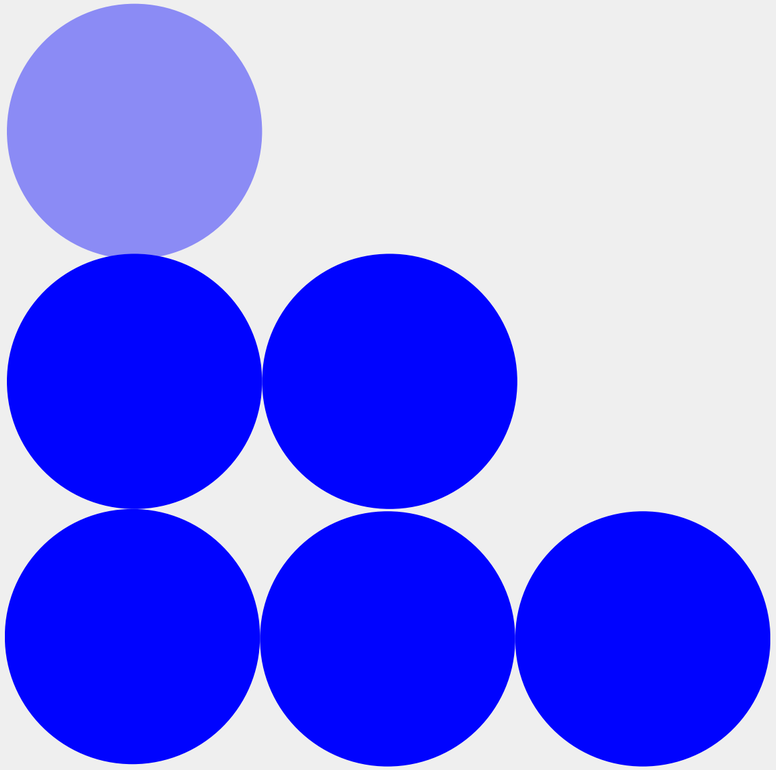
Final Portfolio Project



Ambareen Baig

March 17 2025

Project Idea & Rationale



Problem Statement

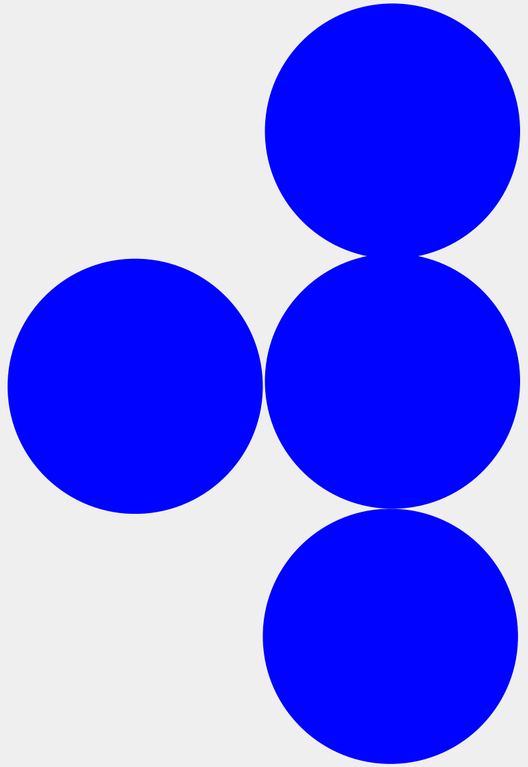
- Manual startup application screening is time-consuming, inconsistent, and prone to bias.
- Larger pool of businesses applying to incubation and acceleration programs, which can be taxing for program teams to sift through while shortlisting for interviews
- Selection board decisions lack predictive insights, leading to potential misjudgments.
- Growing need for an AI-driven, data-backed evaluation system.

Project Objective

- Automate and standardize the startup screening process using AI.
- Reduce work load by using AI to support initial shortlisting for screening interviews
- Improve efficiency, fairness, and accuracy in selection decisions.
- Provide data-driven recommendations to reduce human subjectivity.

Workflow & System Architecture

How Quintessence AI Works



01

Data Consolidation & Preprocessing

- Clean and prepare startup application data.
- Feature engineering to create selection likelihood scoring.

02

AI-Powered Screening

- Random Forest Classifier for structured scoring and classification.
- GPT-3.5 LLM Assistant for qualitative insights and trend analysis.

03

Deployment & User Interface

- Streamlit Dashboard for interactive selection evaluation.
- AI Assistant for answering evaluator queries.

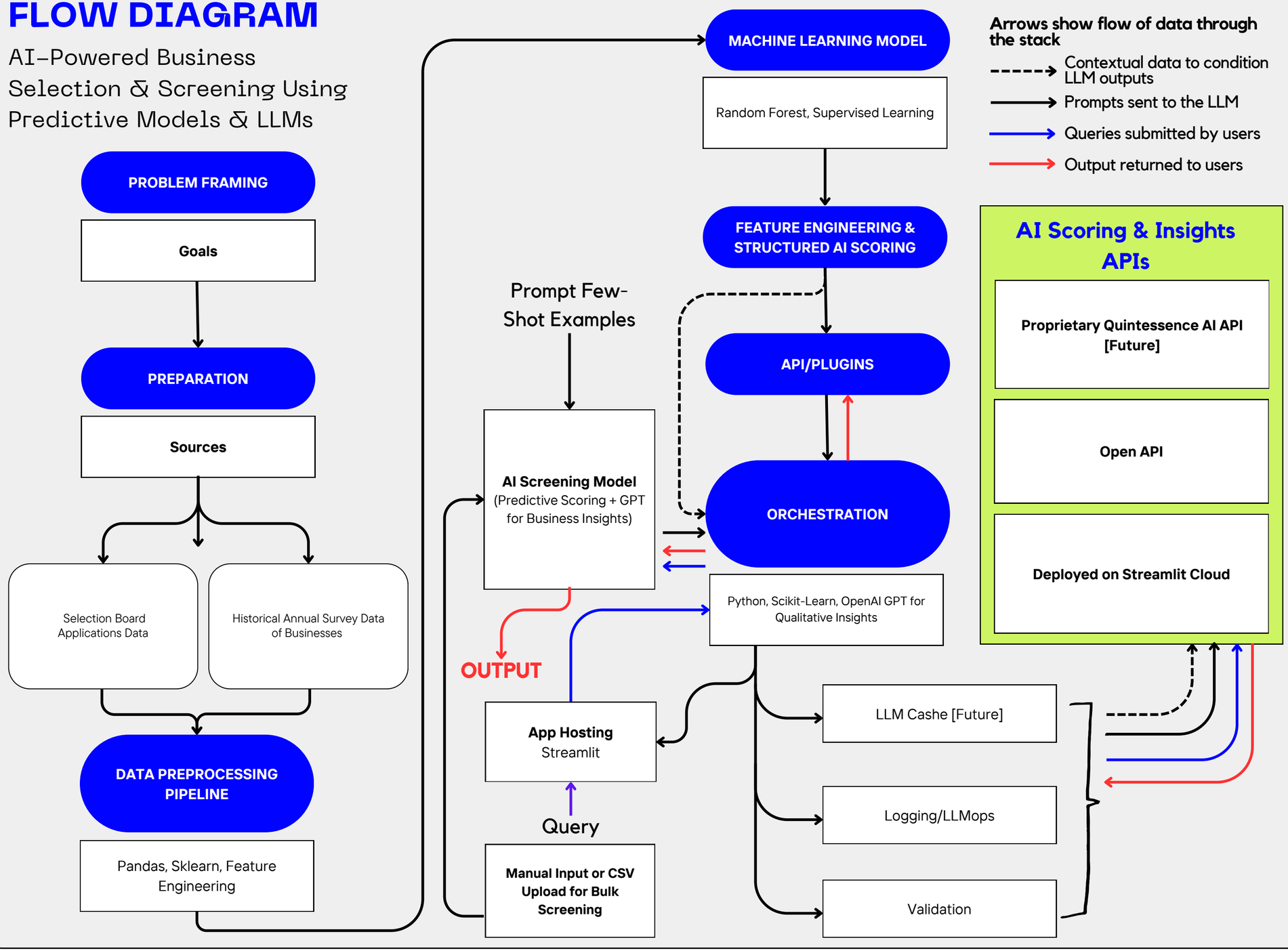
Entity Relationship Diagram

ERD

- **Startup Applications Table:** Stores application details, industry classification, selection outcomes.
- **Selection Scores Table:** AI-generated scores based on structured evaluation.
- **AI Insights Table:** Stores qualitative analysis from LLM-based processing.

FLOW DIAGRAM

AI-Powered Business Selection & Screening Using Predictive Models & LLMs



Technical Details & Methodology

Key Technologies Used

- **Python** – Data Processing & Model Development.
- **Scikit-learn** – Predictive Modeling (Random Forest Classifier).
- **OpenAI GPT-3.5 API** – NLP-based qualitative analysis.
  - **Note**: The OpenAI GPT-powered qualitative chatbot feature was added but is currently experiencing issues post-deployment. Due to time constraints, it is submitted as-is to demonstrate proof of work, but troubleshooting is ongoing.
- **Pandas & NumPy** – Data preprocessing & feature engineering.
- **Streamlit** – Web-based interactive application.
- **SQLite** – Database for structured data storage.

Model Training & Performance Evaluation

Datasets Used

Training Approach

VA Selection Board Data	Annual Survey Data	Annual Survey Data	VA Selection Board Data
Contains recent application data from businesses applying for programs, including their screening status as per evaluator decisions.	A much larger dataset (4,000+ entries) consisting of past program beneficiaries. This dataset was used to compute predicted scores based on the selection criteria and rubric.	The Annual Survey Data was used to train the model on score calculations. **Train-Test Split:** 80% Training | 20% Testing. (This was performed on the Annual Survey Data.)	The VA Selection Board Data was then scored based on selection criteria, and AI-predicted outcomes were compared with human evaluation decisions.

Category	Subcriteria	Scoring Logic	Maximum Points	Total
Market & Growth Potential	Revenue	1.5 pts: Revenue > 500,000; 1.0 pts: 100,000 - 500,000; 0.5 pts: <100,000	1.5	5
Market & Growth Potential	Jobs Created	1.0 pts: >8 jobs; 0.75 pts: 4-8 jobs; 0.5 pts: ,â§3 jobs	1	
Market & Growth Potential	Investment Received	1.0 pts: Investment > 50,000; 0.75 pts: 10,000 - 50,000; 0.5 pts: <10,000	1	
Market & Growth Potential	Number of Clients	1.0 pts: >5000 clients; 0.75 pts: 401-5000; 0.5 pts: 51-400; 0.25 pts: ,â§50	0.75	
Market & Growth Potential	Rural Producers Supported	1.0 pts: >50 rural producers; 0.75 pts: 6-50; 0.5 pts: 1-5; 0.25 pts: 0	0.75	
Team & Expertise	Education Level	1.75 pts: PhD, Masters, Bachelors; 1.25 pts: High School; 0.75 pts: No formal education	1.75	4.25
Team & Expertise	Founder Age	1.5 pts: Founder age <35; 1.0 pts: 35-50; 0.5 pts: >50	1.5	
Team & Expertise	Founder Gender	1.0 pts: Female Founder; 0.5 pts: Male Founder	1	
Value Proposition	Sector Relevance	5.0 pts: Healthcare, ICT, Finance, Agriculture; 4.0 pts: Processed Food; 3.0 pts: Fashion, Construction, Transport, Entrepreneurship; 2.0 pts: Wholesale, Mining, Public Admin, Other	5	5

Performance Metrics

Overall Model Performance: AI vs. VA Selection Board

	Performance
Accuracy	57.61%
Precision (Class 0 - Rejected)	69%
Recall (Class 0 - Rejected)	68%
Precision (Class 1 - Accepted)	33%
Recall (Class 1 - Accepted)	34%
F1-Score (Overall)	58%

Challenges & Solutions

Major Challenges Faced

01

AI was too strict in rejecting startups

No accepted businesses were correctly classified.

02

AI struggled with qualitative aspects

Business potential is not just numbers; human evaluators consider contextual factors.

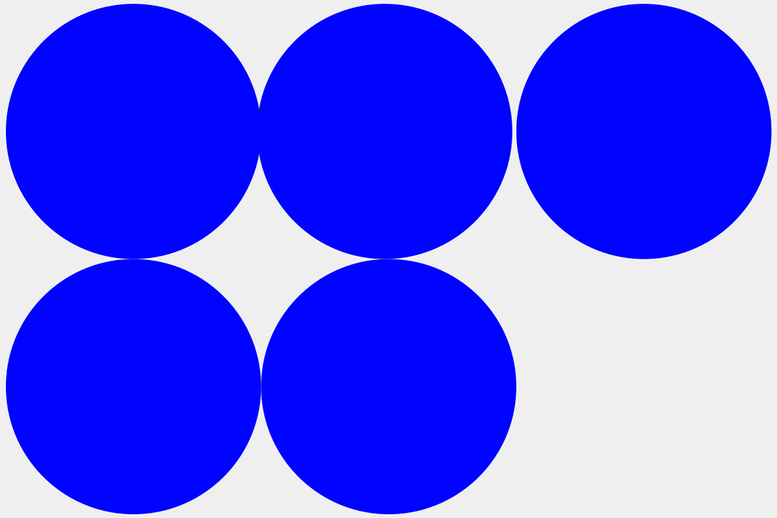
03

High False Positive Rate

AI accepted businesses that the human selection board had rejected.

Challenges & Solutions

Solutions Implemented



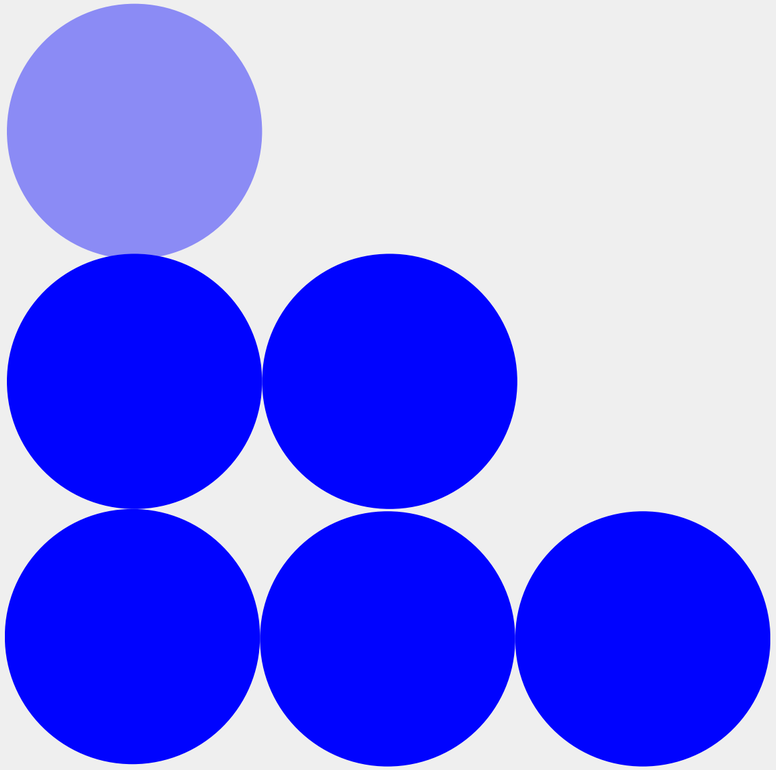
Lowered AI Threshold from 75% to 70% to improve recall.

Introduced a Manual Review category for borderline cases (65%-75%).

Sector-Specific Weighting Adjustments to prioritize high-impact industries.

Refinement of AI Model using more diverse training data.

Project Outcomes & Future Enhancements



Key Achievements

- Developed a working AI-powered screening tool that automates startup evaluations.
- Successfully deployed an interactive dashboard for real-time screening.
- Improved transparency in selection decision-making using AI-powered insights.

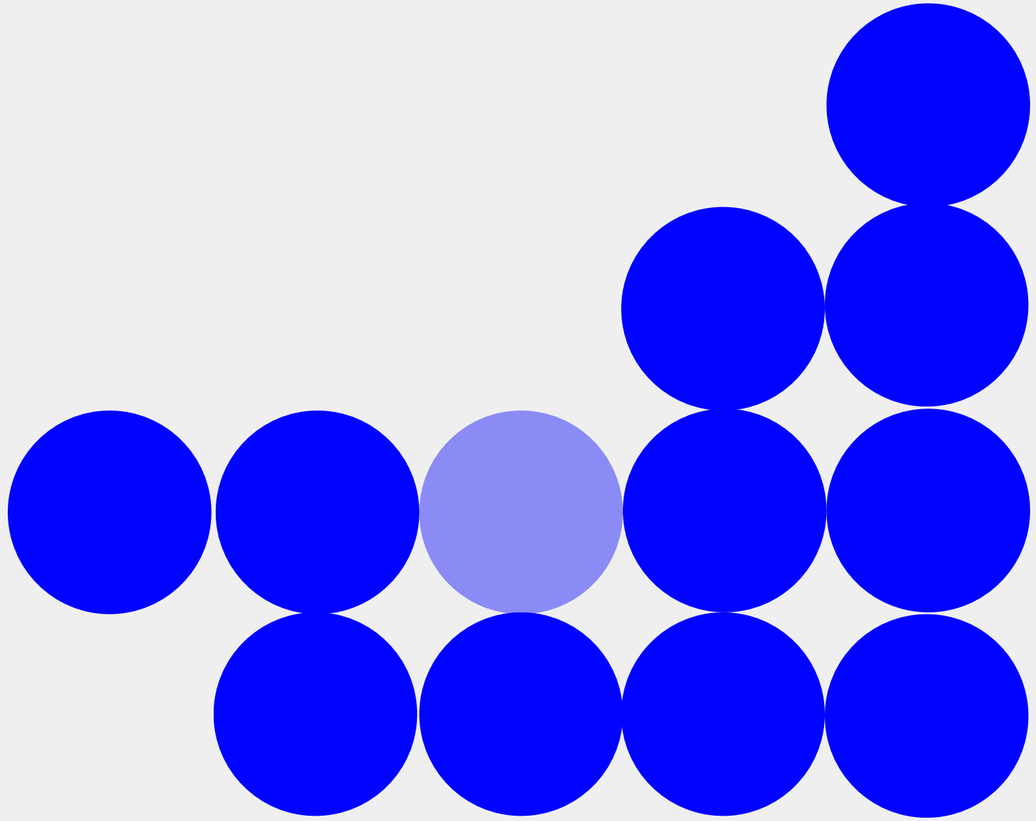
Next Steps & Improvements

- **Investment Selection Expansion**: Refine AI models to assess funding potential, not just selection.
- **Sector-Based Adjustments**: Fine-tune AI for high-impact industries like Healthcare & ICT.
- **Improved NLP Capabilities**: Enhance GPT-3.5 integration for more contextual business analysis.
- **Continuous Learning**: Enable the model to adapt dynamically as more data becomes available.

Launch Quintessence AI

Conclusion

- Quintessence AI represents a step forward in AI-driven startup evaluation in the Pakistan market.
- It improves efficiency, fairness, and consistency in decision-making.
- Further optimizations will make AI more aligned with human evaluators.



GitHub Repo Access Here

Portfolio Project Report