THEME & PROBLEM STATEMENT

Theme: FinTech Innovation

Problem: Hidden Risks in Financial Documents
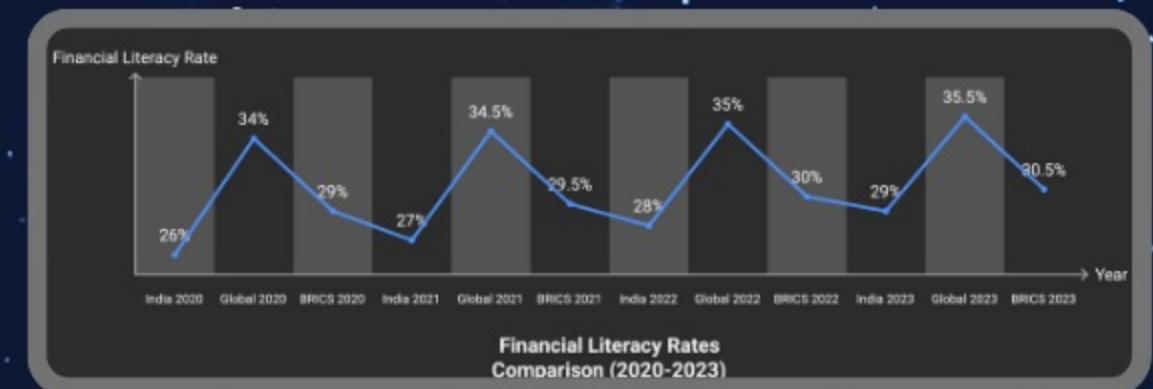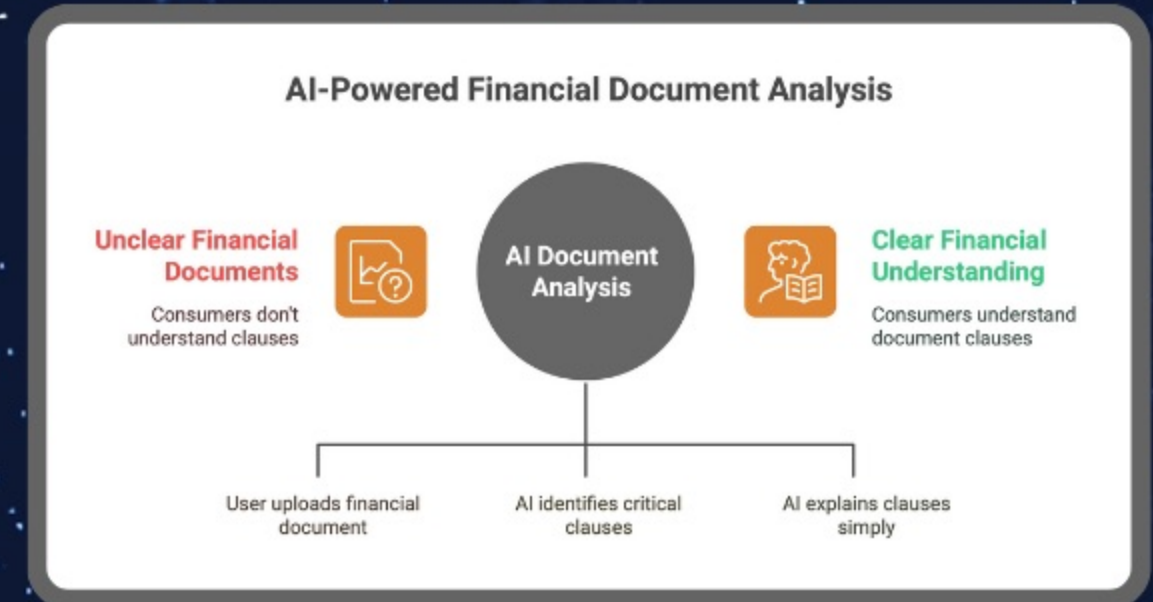
The Problem:

Millions of consumers sign financial documents, such as insurance policies, loans, and credit cards, without fully comprehending intricate clauses. This lack of financial literacy leads to unforeseen penalties, unfavorable terms, and financial losses.

The Challenge:

The accessibility of legal and financial jargon poses a significant challenge for the average individual. Even digitally proficient users often lack the tools to extract valuable insights from complex documents.

Opportunity:

Develop an AI-powered platform that meticulously analyzes and elucidates critical, concealed clauses within real-world documents.

AI-Powered Financial Document Analysis

Unclear Financial Documents — Consumers don't understand clauses

AI Document Analysis

Clear Financial Understanding — Consumers understand document clauses

User uploads financial document | AI identifies critical clauses | AI explains clauses simply

Financial Literacy Rate

26% India 2020 | 34% Global 2020 | 29% BRICS 2020 | 27% India 2021 | 34.5% Global 2021 | 29.5% BRICS 2021 | 28% India 2022 | 35% Global 2022 | 30% BRICS 2022 | 29% India 2023 | 35.5% Global 2023 | 30.5% BRICS 2023

Financial Literacy Rates Comparison (2020-2023)

Despite global improvements, India's financial literacy remains below the curve. The gap ClauseWise AI aims to bridge.
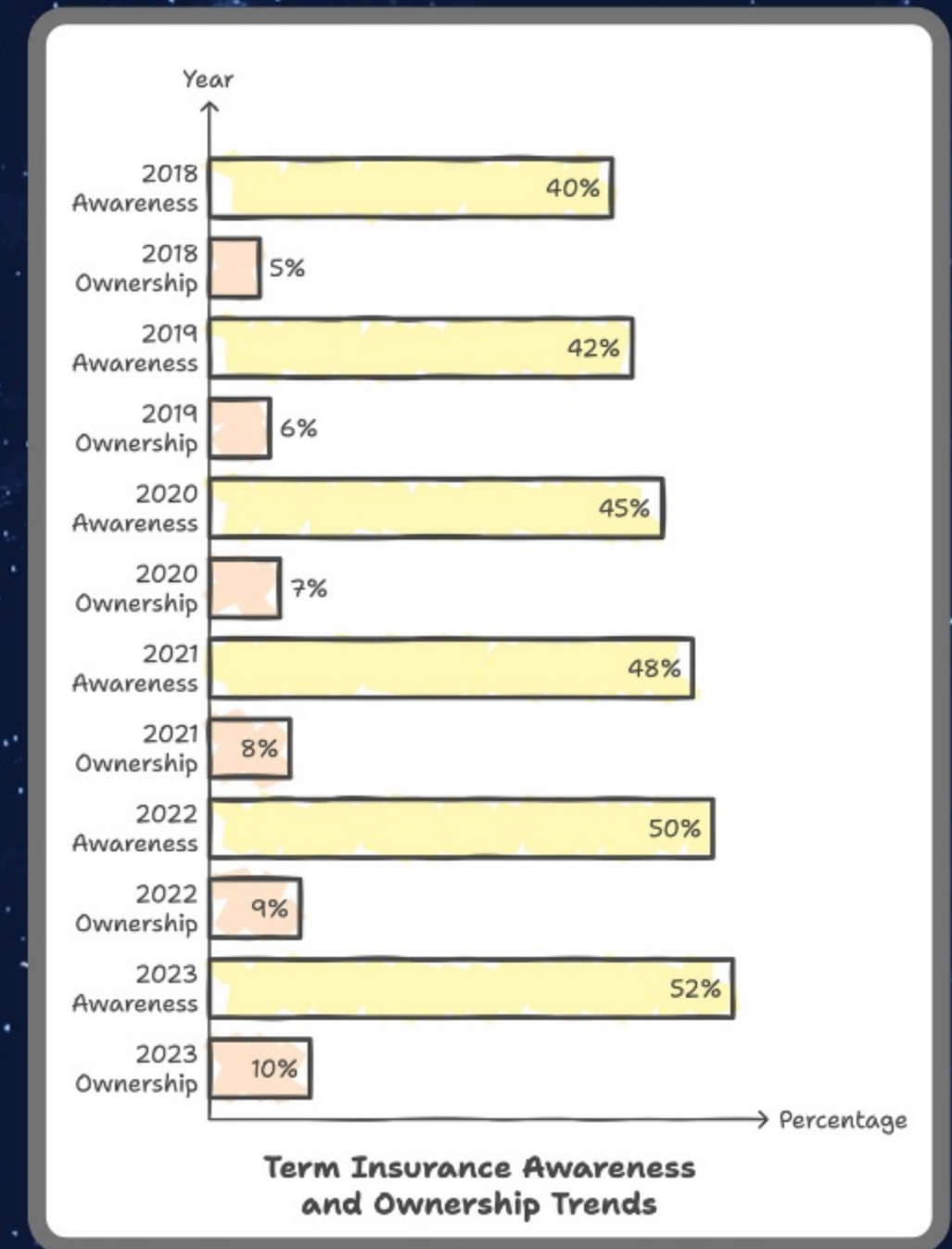
PROPOSED SOLUTION

What It Is:

A smart document analysis platform that helps users understand and converse with their financial documents using AI.

Core Value Proposition:

- Detects hidden clauses and potential risks
- Generates executive summaries of documents
- Offers an interactive AI chat interface for document-based queries
- Teaches financial literacy with a gamified 30-day course
- Works with PDFs, DOCs, and images (OCR-powered)

Impact:

Empowers users to make informed financial decisions, avoid costly mistakes, and develop financial confidence.

Year

	Percentage
2018 Awareness	40%
2018 Ownership	5%
2019 Awareness	42%
2019 Ownership	6%
2020 Awareness	45%
2020 Ownership	7%
2021 Awareness	48%
2021 Ownership	8%
2022 Awareness	50%
2022 Ownership	9%
2023 Awareness	52%
2023 Ownership	10%

Term Insurance Awareness and Ownership Trends

FLOWCHART / DIAGRAM

The ClauseWise AI system follows a streamlined, end-to-end pipeline combining OCR, natural language processing, and AI for seamless financial document analysis:

Step-by-step Breakdown:

Document Upload:

- Users upload files in PDF, DOC, or image formats via drag-and-drop or click-to-browse. Files are validated for size and type.
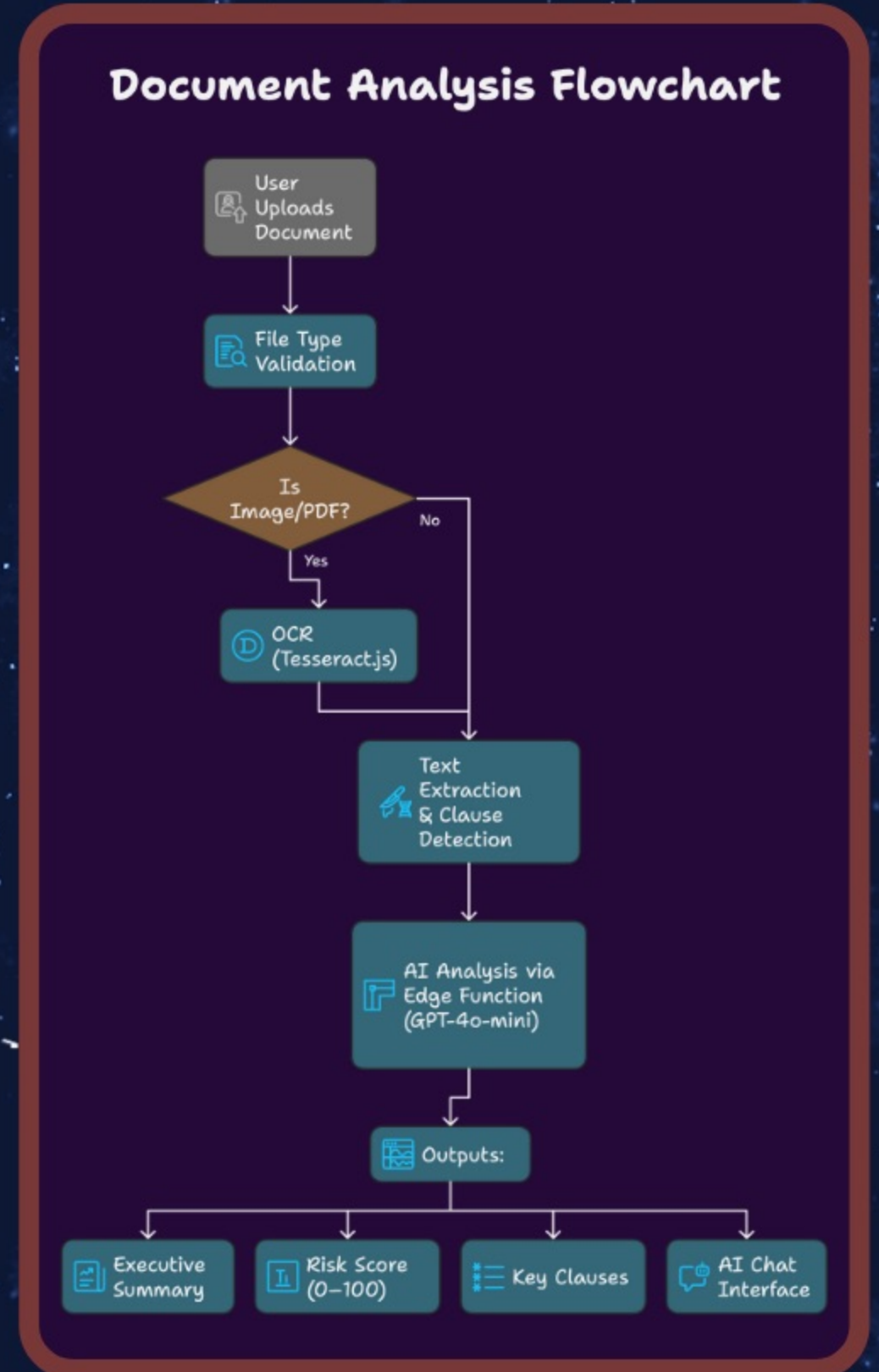
Text Extraction & OCR:

- For image-based documents, client-side OCR (Tesseract.js) extracts text with multi-language support.
- Confidence scoring ensures extraction quality.

Clause Detection & Structuring:

- Extracted text is processed using regular expressions and NLP to identify key sections: terms, fees, hidden clauses, and disclaimers.

AI-Powered Analysis:

- The text is sent to a Edge Function, which calls the GPT-4o-mini model.
- AI returns a structured Result

Document Analysis Flowchart

User Uploads Document → File Type Validation → Is Image/PDF?

- Yes → OCR (Tesseract.js)
- No

→ Text Extraction & Clause Detection → AI Analysis via Edge Function (GPT-4o-mini) → Outputs:

- Executive Summary
- Risk Score (0–100)
- Key Clauses
- AI Chat Interface
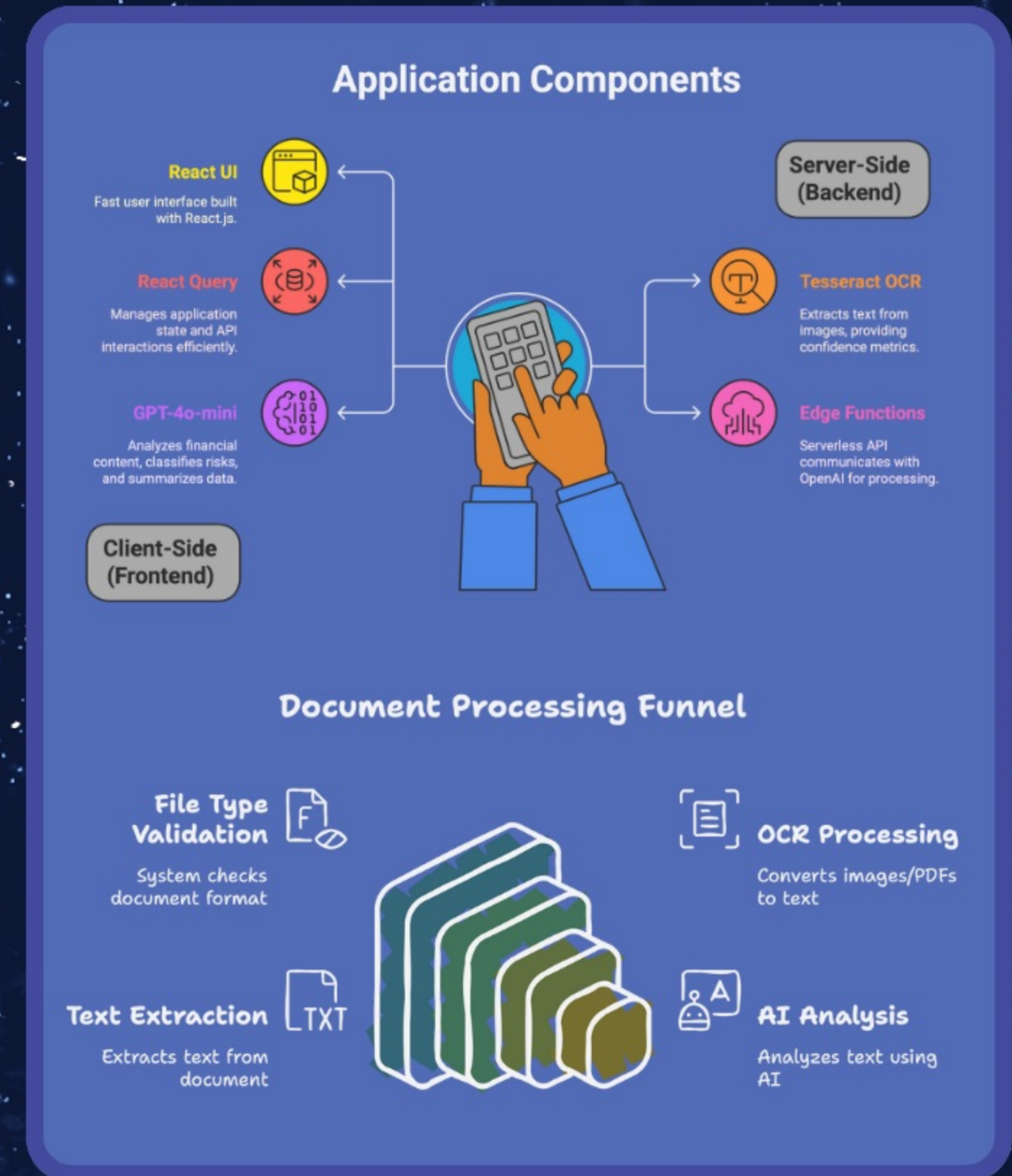
FLOWCHART / DIAGRAM

Client-Side (Frontend):

- React + Vite + Tailwind: Fast, modern UI
- Tesseract.js OCR: Extracts text from scanned images with confidence metrics
- React Query: Efficient state & API management

Server-Side (Backend):

- Edge Functions: Acts as serverless API to communicate with OpenAI
- GPT-4o-mini: Trained prompts analyze financial content, classify risks, and output structured summaries

Data Output:

- JSON-based structured summary
- Risk score
- Highlighted clauses
- Chat-ready document context

Application Components

React UI — Fast user interface built with React.js.

Server-Side (Backend)

React Query — Manages application state and API interactions efficiently.

Tesseract OCR — Extracts text from images, providing confidence metrics.

GPT-4o-mini — Analyzes financial content, classifies risks, and summarizes data.

Edge Functions — Serverless API communicates with OpenAI for processing.

Client-Side (Frontend)

Document Processing Funnel

File Type Validation — System checks document format

OCR Processing — Converts images/PDFs to text

Text Extraction — Extracts text from document

AI Analysis — Analyzes text using AI

FEATURES AND NOVELTY

Key Features:

- **Multi-Format Upload Support:** PDF, DOC, and image uploads are supported.
- **Clause Detection Engine:** Employs pattern recognition and natural language processing (NLP) for accurate clause identification.
- **Risk Scoring System:** Utilizes artificial intelligence (AI) to assess risk levels and provide color-coded risk scores (Low, Medium, High).
- **AI Chat Interface:** Facilitates the interaction with users to inquire about their documents.
- **30-Day Financial Literacy Course:** Offers educational content that can be integrated into daily usage.
- **OCR Processing:** Provides client-side, multilingual OCR capabilities.
- **Executive Summary Generator:** Generates concise and insightful executive summaries.

Novelty:

- Integrates OCR, NLP, and LLM into a unified workflow.
- User-centric, explainable AI design principles.
- Educational layer seamlessly integrated into utility applications.
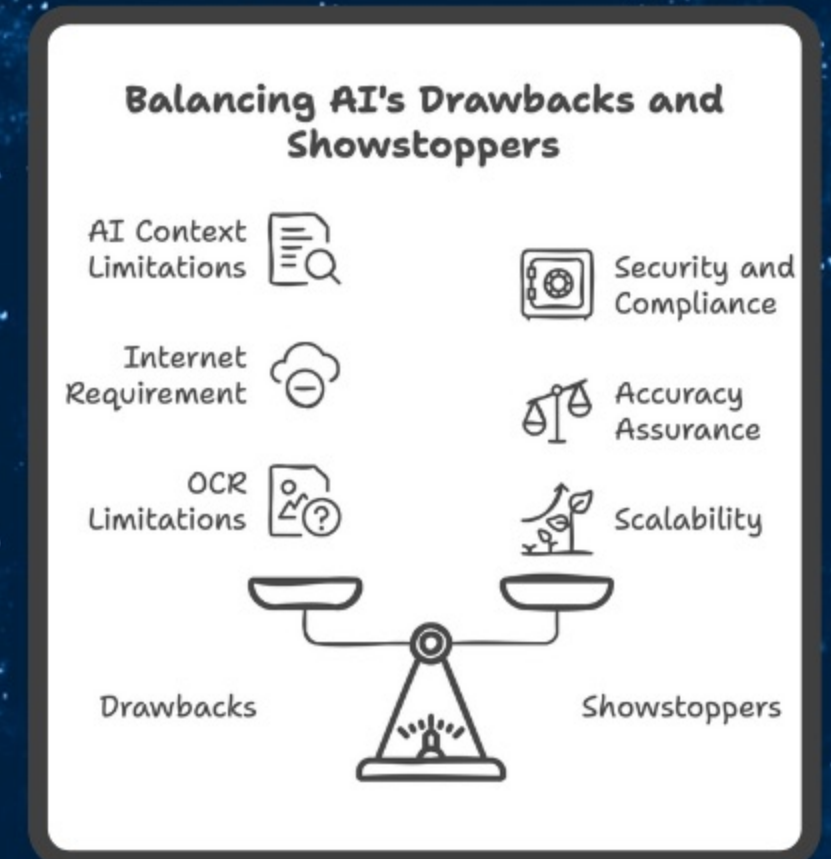- Serverless architecture with high-performance Edge Functions.

Key Features

Upload Support
Supports PDF, DOC, and image uploads for versatility.

Risk Scoring
AI assesses risk levels and provides color-coded scores.

Literacy Course
Offers educational content integrated into daily usage.

Summary Generator
Generates concise and insightful executive summaries.

Clause Detection
Uses pattern recognition and NLP for clause identification.

AI Chat
Allows users to inquire about their documents.

OCR Processing
Provides client-side, multilingual OCR capabilities.

Novelty Features

Unified Workflow
Integrates OCR, NLP, and LLM into a unified workflow.

Explainable AI
User-centric, explainable AI design principles are used.

Integrated Education
Educational layer seamlessly integrated into utility applications.

Serverless Architecture
Serverless architecture with high-performance Edge Functions is used.

DRAWBACK AND SHOWSTOPPERS

Drawbacks / Limitations:

- **AI Context Limitations:** GPT-4o-mini may truncate lengthy documents during analysis.
- **Internet Requirement:** Cloud-based functions necessitate internet connectivity.
- **OCR Limitations:** Inadequate image quality compromises accuracy.
- **Privacy Concerns:** Users may exhibit reluctance to upload sensitive documents.
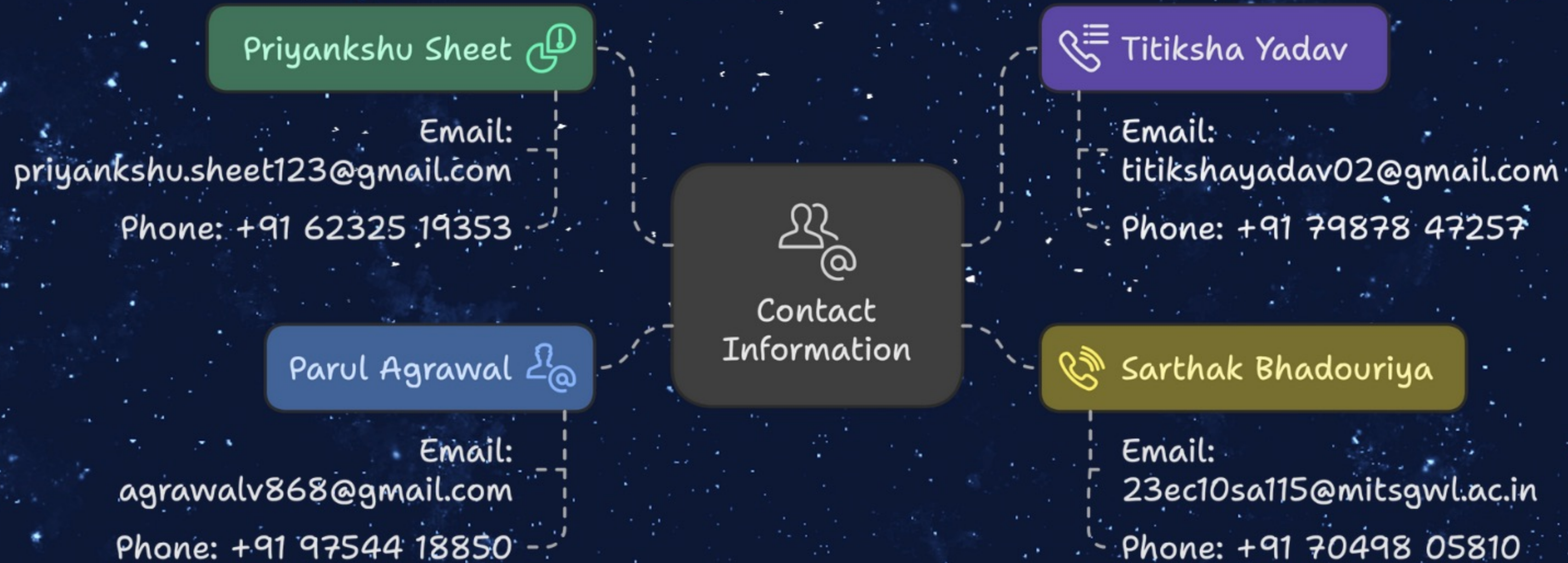
Showstoppers:

- **Security and Compliance:** Requires end-to-end encryption and user trust.
- **Accuracy Assurance:** AI must refrain from generating misleading summaries or omitting crucial clauses.
- **Scalability:** Serverless infrastructure must efficiently manage large document processing at scale.

Balancing AI's Drawbacks and Showstoppers

AI Context Limitations
Security and Compliance
Internet Requirement
Accuracy Assurance
OCR Limitations
Scalability

Drawbacks | Showstoppers

Balancing innovation with limitations to build a scalable, secure AI solution.

TEAM CLAUSEWISE

Contact Information

Priyankshu Sheet

Email:
priyankshu.sheet123@gmail.com

Phone: +91 62325 19353

Parul Agrawal

Email:
agrawalv868@gmail.com

Phone: +91 97544 18850

Contact Information

Titiksha Yadav

Email:
titikshayadav02@gmail.com

Phone: +91 79878 47257

Sarthak Bhadouriya

Email:
23ec10sa115@mitsgwl.ac.in

Phone: +91 70498 05810