

# SafePayAI - Semi-Technical Pitch

*"Every 8 seconds, someone in India falls victim to a UPI fraud. As digital payments skyrocket to billions of transactions daily, fraudsters are evolving faster than traditional rule-based detection systems can keep up. SafePayAI changes that game."*

---

## The Problem

---

Traditional fraud detection relies on **static rules** — if amount > ₹50,000, flag it. But fraudsters adapt instantly.

### Core challenges:

- **Class imbalance:** Legitimate transactions outnumber fraudulent ones 1000:1
- **Evolving patterns:** Fraud techniques change weekly
- **False positives:** Blocking legitimate users damages trust
- **Real-time demands:** Decisions must happen in milliseconds

## Our Solution

---

SafePayAI uses a **two-stage AI architecture**:

### 1 GAN-Powered Data Augmentation

We use **Generative Adversarial Networks** to solve the class imbalance problem:

- The **Generator** creates synthetic transaction data from random noise
- The **Discriminator** learns to distinguish real vs synthetic data
- This adversarial training produces high-quality synthetic fraud data

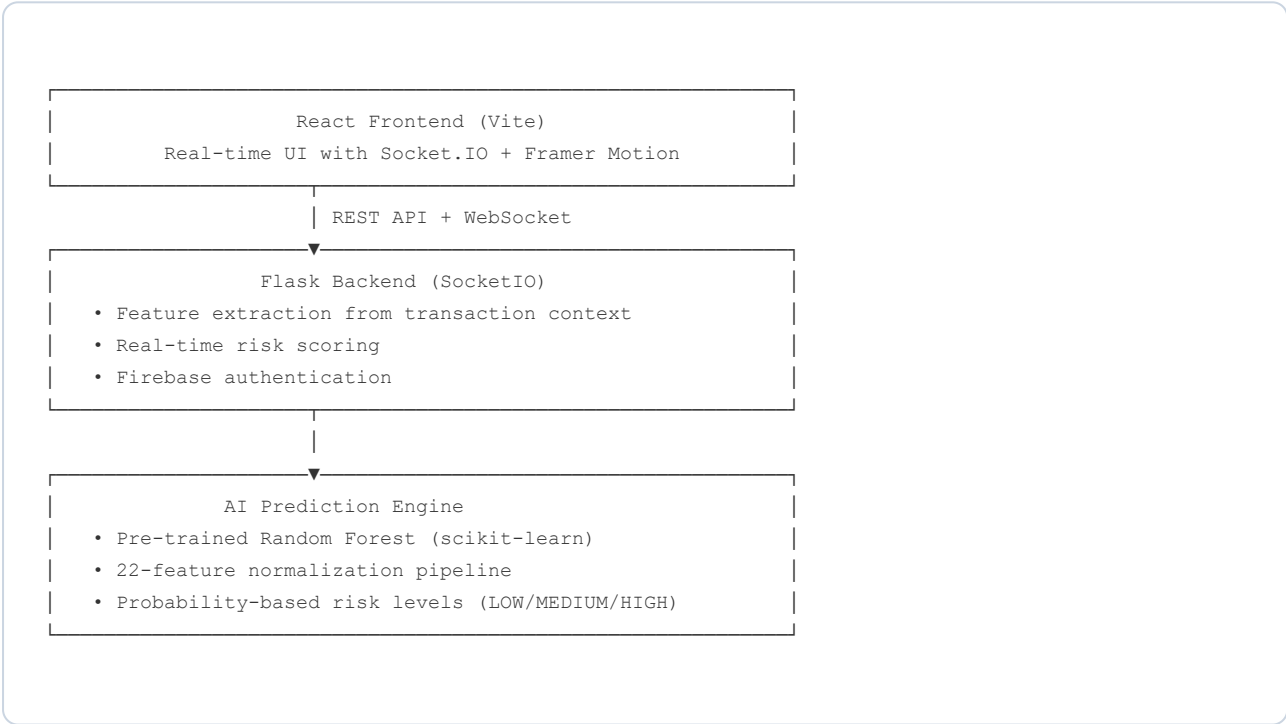
**Result:** Our model learns from 10x more fraud examples without collecting more real fraud cases.

### 2 Random Forest Classifier

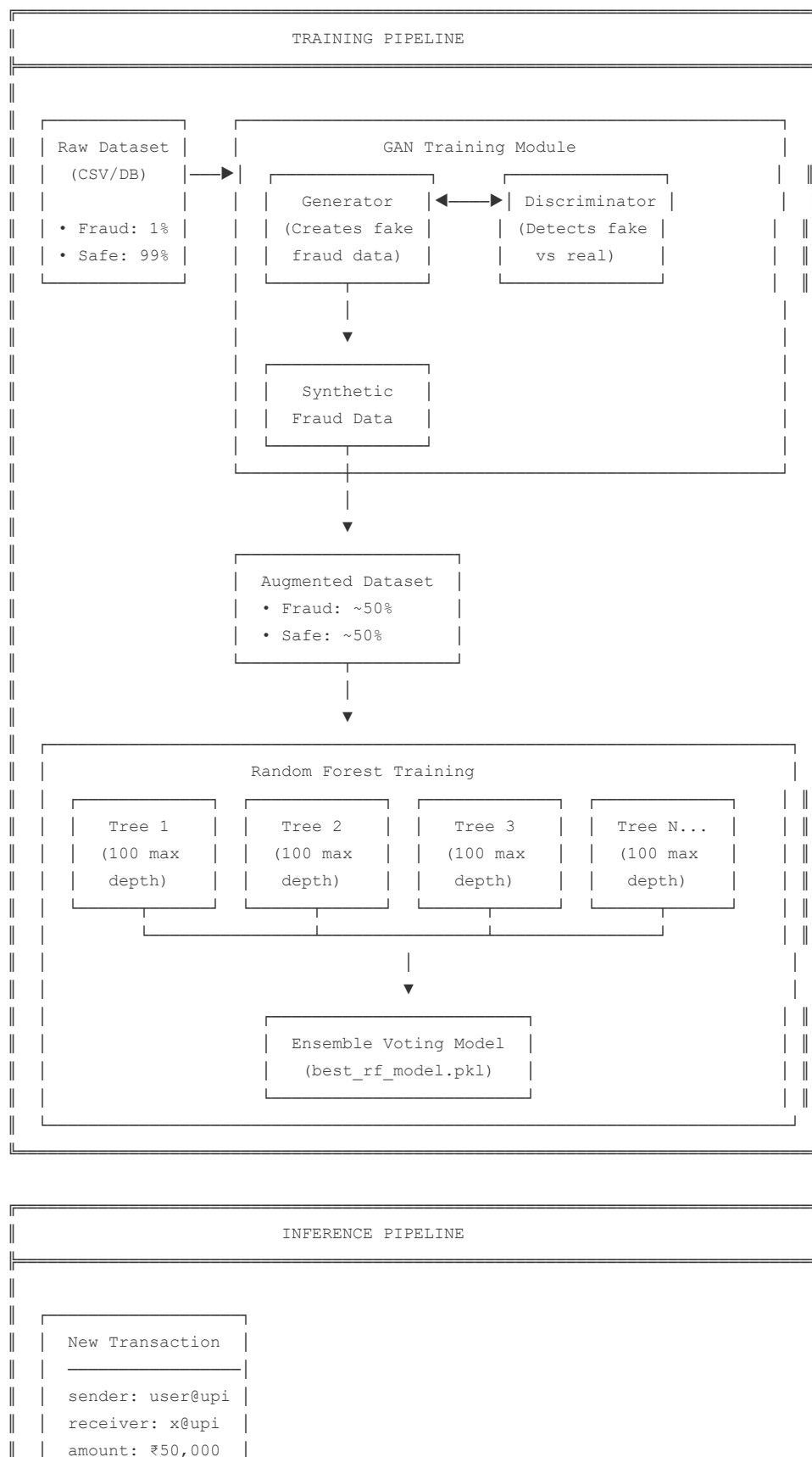
Our production model analyzes **22 real-time features** per transaction:

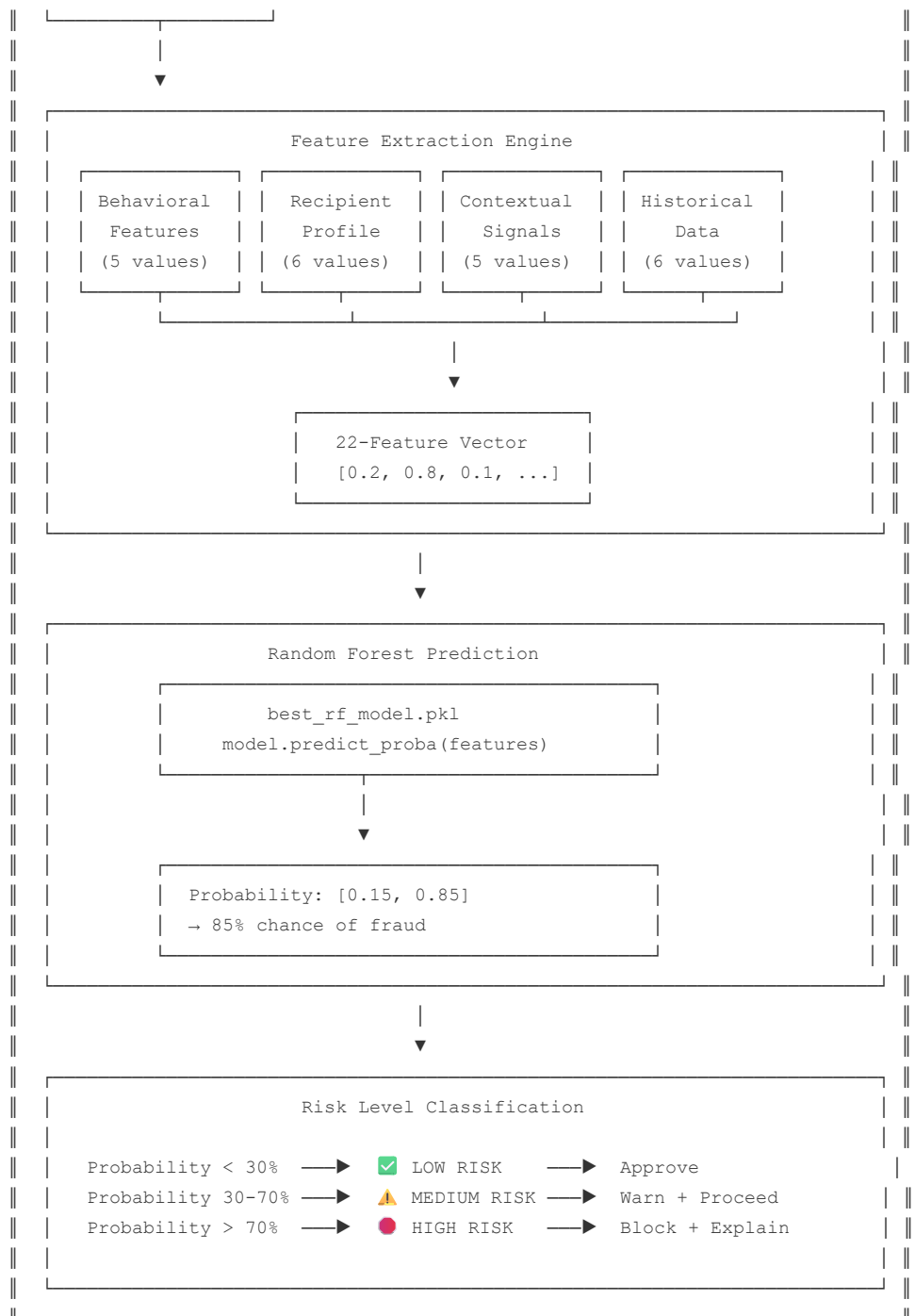
Feature Category	Examples
Behavioral	Transaction frequency, amount deviation from history
Recipient Risk	Trust score, verification status, blacklist status
Contextual	Transaction hour, geo-location anomalies
Device	Device fingerprint, VPN/proxy detection
Historical	Past fraud complaints, account age

## Technical Architecture



## ML Pipeline Architecture





Key Differentiators

What Others Do	What We Do
Rule-based thresholds	ML-driven pattern recognition
Block & frustrate users	Risk levels with explanations
Batch processing	Real-time prediction (<100ms)
Fixed fraud patterns	GAN-generated evolving patterns
Binary yes/no	Probability scores + risk factors

Demo Highlights

Scenario	Recipient	Result
✔ Send to trusted merchant	trusted.merchant@upi	LOW risk, instant approval
⚠ Send to suspicious account	suspicious.account@upi	WARNING with risk factors displayed
🔴 Send to known fraud actor	fraud.actor@upi	BLOCKED with explanation

"The user sees WHY their transaction is flagged — building trust, not frustration."

Impact & Metrics

<div>Model Accuracy</div> <div>~94%</div> <div>on test data</div>	<div>False Positive Rate</div> <div>&lt;5%</div> <div>minimal user friction</div>
<div>Inference Time</div> <div>&lt;50ms</div> <div>per transaction</div>	<div>Features Analyzed</div> <div>22</div> <div>real-time signals</div>

## Vision

Integrate SafePayAI as a **middleware layer** for any payment platform — protecting millions of users with AI-powered, explainable fraud detection.

## Q&A Preparation

Potential Question	Answer
<i>Why Random Forest over Deep Learning?</i>	Interpretability + speed. We can explain which features triggered a flag, and inference is instant without GPU.
<i>How does the GAN help?</i>	It synthetically generates realistic fraud scenarios, solving the 1:1000 class imbalance without overfitting.
<i>What about privacy?</i>	All predictions happen locally on transaction metadata — no card/account numbers are stored or transmitted.
<i>How scalable is this?</i>	Flask + pre-loaded model can handle thousands of requests/sec. For production, we'd add Redis caching and load balancing.

**Closing Statement:** *"SafePayAI doesn't just detect fraud — it **anticipates** it. By combining generative AI for training and ensemble models for prediction, we've built a system that evolves as fast as the fraudsters do. The future of payment security isn't just reactive. It's predictive."*

## Research References

1. Zong, K. et al. (2025). *Detection of AI Deepfake and Fraud in Online Payments Using GAN-Based Models*. arXiv:2501.07033. [arxiv.org/abs/2501.07033](https://arxiv.org/abs/2501.07033)
2. Gao, Y. et al. (2023). *Advancing Financial Fraud Detection: Self-Attention Generative Adversarial Networks*. ScienceDirect. [sciencedirect.com/science/article/abs/pii/S1544612323012151](https://sciencedirect.com/science/article/abs/pii/S1544612323012151)
3. Taha, A. A. & Malebary, S. J. (2023). *A Survey on GAN Techniques for Data Augmentation in Credit Card Fraud Detection*. MDPI. [mdpi.com/2504-4990/5/1/19](https://mdpi.com/2504-4990/5/1/19)
4. Zhang, X. et al. (2024). *Utilizing GANs for Fraud Detection: Model Training with Synthetic Transaction Data*. arXiv:2402.09830. [arxiv.org/abs/2402.09830](https://arxiv.org/abs/2402.09830)

5. Chen, Y. et al. (2025). *Semi-Supervised Bayesian GANs with Log-Signatures for Uncertainty-Aware Credit Card Fraud Detection*. MDPI Mathematics. [mdpi.com/2227-7390/13/19/3229](https://doi.org/10.3390/math13193229)
6. Kumar, R. et al. (2025). *Optimizing Credit Card Fraud Detection with Random Forests and SMOTE*. Nature Scientific Reports. [nature.com/articles/s41598-025-00873-y](https://doi.org/10.1038/s41598-025-00873-y)
7. Bhattacharyya, S. et al. (2018). *Ensemble Learning for Credit Card Fraud Detection*. ACM CODS-COMAD. [dl.acm.org/doi/10.1145/3152494.3156815](https://dl.acm.org/doi/10.1145/3152494.3156815)
8. Lee, J. et al. (2024). *An Integrated Multistage Ensemble Machine Learning Model for Fraudulent Transaction Detection*. Journal of Big Data. [journalofbigdata.springeropen.com](https://doi.org/10.1007/s13201-024-01568-1)