Investor EdTech App for India — Spec, Architecture & Roadmap (SEBI-inspired)

1) Problem & Goals

Problem. Retail investors struggle with market basics, risk assessment, and discerning credible advice, especially in vernacular languages. **Goal.** Build a safe, engaging learning app that:

- Teaches foundations (markets, products, risks, scams), algo/HFT concepts (responsibly), and portfolio diversification.
- Uses ≥**15-min delayed market data** + **paper trading** to learn by doing.
- Offers **vernacular translation + summarization** of authoritative sources (SEBI/NISM/exchanges) with clear citations and disclaimers.
- Tracks progression, awards badges, and nudges healthy investor behavior.

Key constraints. Safety/compliance first; low-end device performance; offline-first; minimal data costs.

---

2) Target Users & Personas

- **First-time investor (18–30)**: Learning basics, small savings, mobile-first, Hindi/Tamil/Bengali, wants confidence.
- **Returning investor (25–45)**: Has losses from tips/YouTube; wants structured learning + practice.
- **Regional educator/NGO**: Needs classroom mode, printable sheets, and trackable cohorts.

---

3) Solution Concept (Two Tracks)

Track A — Learn + Sim Trade

Interactive tutorials → scenario drills → quizzes → risk profile → **virtual portfolio** with delayed data. AI coach explains P&L swings, diversification, and risk.

Track B — Vernacular Summarizer

In-app **translate + summarize** trusted articles/circulars into Indian languages (Hindi, Bengali, Marathi, Tamil, Telugu, Kannada, Gujarati, Malayalam, Odia, Punjabi). Provides bullet summaries, key terms, and glossary cards; keeps links to originals.

---

4) Core Features

1. **Guided curriculum**: Markets 101, Instruments (Equity, MF/ETF, Bonds), Risk/Return, Frauds/Red flags, Taxes basics, Algo/HFT concepts (history, market microstructure, latency, regulation, risks), Diversification.
2. **Risk Profiler**: Questionnaire → risk bucket (Conservative/Moderate/Aggressive) + explanation; used to tailor content + practice.
3. **Virtual Trading (Paper Trading)**: Create watchlists; place market/limit orders at delayed prices; T+2 settlement emulation; portfolio/sector drift view; realistic constraints (slippage bands, no intraday leverage by default); trade journal with lesson prompts.
4. **Delayed Market Data**: ≥15-min delay; daily OHLCV; index/ETF baskets; corporate actions (split/bonus) simulated.
5. **Algo/HFT Classroom** (responsible): Simulated strategies on **delayed/historical** data (moving average crossover, mean reversion, momentum); **no real-time signals**; heavy emphasis on risks and why retail should avoid live HFT.
6. **Vernacular Translate+Summarize**: Chunk documents, glossary mapping, output graded by reading level; side-by-side compare EN↔local.
7. **Scam Shield**: Interactive red-flag lab (messages, screenshots); user identifies pump-and-dump, unsolicited tips, get-rich-quick.
8. **Assessments & Badges**: MCQs, scenarios, and mini-projects; badge gating for new features (e.g., paper-trade only after Markets 101 pass).
9. **Coach & Nudges**: Explains variance, urges diversification, shows drawdown probability charts.
10. **Cohort/Classroom Mode**: Facilitator dashboard, assignments, printable PDFs, attendance via QR.
11. **Accessibility & Offline**: Full Hindi UX; optimized for low-RAM phones; downloadable lessons.

---

5) Content Map (Curriculum Outline)

- **Module 1: Foundations** — What is a stock/exchange/depository; order types; bid-ask; settlement; basic math (CAGR, variance).
- **Module 2: Risk & Behavior** — Volatility vs risk, drawdown, diversification, cognitive biases, position sizing, SIPs.
- **Module 3: Products** — Equity, ETFs/MFs, Bonds (G-Secs/SDLs), Derivatives (intro + cautions), IPOs.
- **Module 4: Fraud & Safety** — Unsolicited tips, Telegram/WhatsApp schemes, celebrity impersonation, KYC/UPI hygiene, grievance redress basics.
- **Module 5: Algo/HFT — Concepts** — Matching engines, latency, limit order book (LOB), basic strategies (paper only), risk controls, regulation overview.
- **Module 6: Portfolio** — Asset allocation, rebalancing, sector/cap mix, risk-aligned model portfolios for **education only**.
- **Capstone** — Build a paper portfolio, write an Investment Policy Statement (IPS), and reflect in a journal.

---

6) Learning Progression & Rewards

- XP points per lesson; streaks; bronze/silver/gold badges.

- Skill tree unlocks: complete Risk 101 → unlock paper trading; pass Fraud lab → unlock community.
- Shareable certificate (name, date, module list, score).

---

7) Compliance & Safety Guardrails

- **Disclaimers**: Education only; no investment advice; virtual trading not indicative of future results.
- **Data policy**: Only delayed/historical; clear timestamp labels; no real-time order routing.
- **Feature gating**: Derivatives simulations unlock only after Risk 101; HFT sandbox uses historical snapshots with capped order rates.
- **Content provenance**: Show original source links and last-updated time.
- **Red-flag surfacing**: Before any simulation trade, show risk callouts for concentration, leverage, or penny-stock exposure.

---

8) System Architecture

Client (mobile/web)

- React Native (Android-first), React Web; i18n; offline caching; LOB visualizer; charting (lightweight).

Backend

- API: FastAPI/Node; Auth (email/OAuth/phone OTP); Rate limiting; Content & quiz APIs.
- **Data services**: Market data fetcher (scheduled), corporate-action normalizer, OHLCV store.
- **Simulation engine**: Order book simulator (VWAP/partial fills), P&L engine, rebalancer, journal triggers.
- **NLP/LLM**: Translation + summarization pipelines; glossary extraction; toxicity/claims filters.
- **Storage**: Postgres (users, progress, trades), S3 (assets), Redis (cache), Vector DB (content chunks & glossaries).

Data Flow (simplified)

1. Fetch delayed market data → normalize → store OHLCV.
2. User submits paper order → simulation engine prices at last delayed tick ± slippage → portfolio updated.
3. Lesson opened → content chunks pulled from CMS → optional summarization/translation → cached offline.

---

9) Tech Stack

- **App**: React Native Expo; Zustand/Redux; React Query; Recharts/Victory for charts.
- **Backend**: FastAPI (Python) or Node (NestJS) with Pydantic/TypeORM; Celery/RQ jobs.
- **NLP**: Open-source MT (IndicTrans2), MarianMT fallback; sentencepiece tokenization; cue-glossary mapping.

- **Vector**: FAISS/pgvector.
- **CI/CD**: GitHub Actions; Play Store internal testing tracks.

---

10) Data Model (key tables)

```
-- users
(id, phone_or_email, locale, consent_flags, created_at)

-- progress
(user_id, module_id, lesson_id, score, passed, timestamp)

-- orders_sim
(order_id, user_id, symbol, side, qty, order_type, limit_price, status,
placed_at, filled_qty, avg_fill_price)

-- portfolios
(user_id, symbol, qty, avg_cost, sector, updated_at)

-- prices_ohlcv
(symbol, date_time, o, h, l, c, v, is_delayed, source_ts)

-- content_chunks
(id, source_url, lang, text, embedding, updated_at)

-- glossary
(term, definition_en, definition_local, examples)
```

---

11) Simulation & Risk Design

Order fill logic (simplified):

- Market order: fills at next delayed last-traded price (LTP) with configurable slippage band (e.g., ±0.1–0.5%).
- Limit order: fills if LTP crosses limit; otherwise remains open.
- Partial fills & commissions simulated; circuit breakers respected.

Risk checks:

- Concentration >30% in one stock → warning.
- Penny stock flag; illiquidity check using ADV (simulated with percentile filter).
- Max daily turnover limits by risk bucket.

Rebalancer:

- Target weights by risk bucket; rebalance thresholds (5–10% drift); communicates trades as batch journal entries.

---

12) Quiz/Assessment Engine

- MCQ, scenario-based, and calculation items.
- **Blueprinting**: Each question tagged to LO (learning objective) + difficulty.
- **Adaptive**: Item bank with IRT-style difficulty approximation; serve more practice where weak.
- **Explanations** after answer with references.

---

13) Vernacular Translation & Summarization

Pipeline

1. Crawl/ingest white-listed sources.
2. Chunk by semantic sections; detect code/figures/tables.
3. MT → Post-edit rules: numbers, entities, finance lexicon mapping.
4. Summarize to 3 levels: 60-sec, 5-min, deep dive.
5. Glossary auto-generation; back-translation quality checks; human review for top pages.

UI: Side-by-side EN↔local; tap any term for tooltip; audio TTS for accessibility.

---

14) Accessibility, Performance, Offline

- 2G/3G friendly, <15 MB base APK target; lazy loaded modules.
- Text-first design; charts simplified; downloadable lesson packs.
- Screen-reader labels; dyslexia-friendly font option; high-contrast mode.

---

15) Analytics & KPIs

- **Learning**: Pre/post scores, time-on-task, retention by module.
- **Behavioral**: Paper trades/week, diversification index, journal completions.
- **Safety**: % trades triggering risk flags, scam-recognition accuracy.
- **Engagement**: 7-day retention, completion rate, NPS.

---

16) Rollout Plan (12 Weeks)

Phase 0 (Week 1–2) — Content skeleton (Modules 1–2), Hindi localization, basic quiz engine, delayed data ingest. **Phase 1 (Week 3–5)** — Paper trading MVP (equities + ETFs), risk profiler, portfolio dashboard. **Phase 2 (Week 6–8)** — Vernacular summarizer (Hindi→Bengali/Tamil), glossary, Scam Shield lab. **Phase 3 (Week 9–10)** — Algo/HFT classroom (historical only), strategy sandboxes, journal prompts. **Phase 4 (Week 11–12)** — Classroom/cohort mode, certificates, polish & Play Store beta.

---

17) Risks & Mitigations

- **Misinterpretation as advice** → prominent disclaimers; no individualized stock picks; education-only wording; advice filter in copy.
- **Data costs/performance** → pre-compute summaries; compress assets; server-side rendering; delta updates.
- **Translation drift** → glossary lock, back-translation QC, human-in-loop for top pages.
- **User expectations of HFT** → frame as *understanding*, not *doing*; highlight costs/infrastructure barriers and regulation.

---

18) Sample Screens (Wireframe Descriptions)

1. **Home**: Progress ring, “Continue Learning,” watchlist (delayed), today’s concept card.
2. **Lesson View**: Step cards, video/text toggle, glossary rail, quiz chip.
3. **Trade Ticket**: Symbol search → LTP (delayed badge), order type, qty, risk callouts, confirm sheet with education pointers.
4. **Portfolio**: Holdings, P&L, sector wheel, diversification score, rebalance suggestion.
5. **Summarizer**: Paste