MGT 8803: AI in Business

Organizational Implementation Plan - Revised

Team: Swapna Gadre, Atharva Bhalerao, Abhishek Kalokhe

Success Metrics and Evaluation Approach

To evaluate StonksAI, we focus on both technical accuracy and user experience, ensuring the tool delivers measurable forecasting improvements while remaining explainable and trustworthy.

Evaluation Methodology:

We use standard time-series regression metrics such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) to quantify prediction accuracy. For sentiment analysis, we benchmark directional correctness, i.e., whether the LLM's predicted movement (up/down) matches the actual stock movement in the following month. Our business benchmark is the performance of the S&P 500 or basic index strategies.

We employed iterative testing during model refinement by adjusting prompt templates, updating training windows, and measuring performance differences in both accuracy and interpretability.

Frequency & Responsibility

During development, we evaluate system performance weekly using fresh data inputs. In production, we implement a scheduled update cadence:

- News articles are fetched daily, covering the most recent two-week window to ensure sentiment signals remain timely and relevant.
- Stock market data is retrieved every 30 days, aligning with our LSTM’s monthly prediction frequency.

Responsibilities are divided as follows:

- **Data scientists & ML Engineers**: Monitor model health, drift, and accuracy
- **Product team**: Evaluate user feedback and engagement metrics
- **Compliance lead**: Review fairness, explainability, and prompt alignment. This structure ensures both model and user performance are continuously evaluated.

Ethical, Privacy, and Bias Considerations

StonksAI is built with user trust, data integrity, and ethical AI use at its core. Although our platform currently uses only public data (stock prices and financial news), we proactively address key ethical considerations in line with industry best practices.

Identifying Potential Risks

- **Bias & Fairness**: While we don't collect demographic data, biases can arise from news source framing or LLMs interpreting sentiment disproportionately. For example, headlines focusing on volatility may exaggerate market movement, skewing LLM responses.
- **Privacy & Data Security**: We do not use personalized user data, but in future iterations, e.g., if we personalize predictions, strict adherence to GDPR/CCPA will be essential. No private financial data is shared with third-party models unless securely processed.
- **Ethical Use & Transparency**: Users must understand that StonksAI is an advisory aid, not financial advice. LLM explanations may hallucinate or overstate causal relationships between news and market outcomes.

Mitigation Strategies

- **Technical safeguards** include:
  - Restricting LLM prompts to cite only retrieved news content
  - Using confidence thresholds for ensemble outputs
  - Logging prediction rationale for auditability
- **Policy controls**:
  - Clear disclaimers throughout the UI
  - Opt-in feedback system for flagged explanations
- **User-centric design**:
  - Explanations are always paired with raw data (e.g., original article summary)
  - Users can toggle between LSTM-only and LLM-enhanced forecasts
  - No personalized data is required to use the tool

Monitoring & Remediation

A bias and explanation audit will be conducted quarterly. LLM outputs are stored with article references for manual verification. In case of hallucinated or misleading outputs, we can roll back prompt versions or filter out problematic inputs using keyword-level rules. We also enable user reporting of suspect predictions or explanations.

Organizational Rollout and Practical Challenges

StonksAI is designed for scalable rollout through a streamlined pilot-to-production plan, minimizing disruption while maximizing integration potential across platforms.

Integration and Scalability

The tool is built on a modular Streamlit web interface. For broader integration, we can deploy our forecasting engine via a REST API to be embedded in fintech dashboards or advisor platforms. All components are cloud-compatible and can scale via containerized deployment (Docker on AWS/GCP). Model refresh and retraining pipelines can be automated using scheduling tools like Airflow.

Change Management

We propose launching first to a pilot group of retail investors or university finance clubs, collecting usage and trust feedback. Key stakeholders include:

- **Product managers**: Ensure roadmap alignment and UI clarity
- **Compliance/legal advisors**: Vet model outputs and user disclosures
- **End users**: Validate explainability, usefulness, and usability

Training is minimal due to the interface's simplicity, but workshops or onboarding videos can support institutional deployments. Output logs and embedded tooltips improve transparency and usability.

Resource Allocation and Budgeting

Initial implementation was completed by a team of 3 students. To scale, we estimate needing:

- 1-2 part-time ML Engineers for model maintenance
- Cloud budget of ~$50-$100/month for model serving and logging
- ~$100/month budget for OpenAI and Alpha Vantage APIs

Total rollout cost remains low for an MVP (Minimum Viable Product) with most investment centered on monitoring, data acquisition, UX testing, and documentation for trusted deployment.