

Project Scoping Submission – Echo AI

Team Members:

- **Abhisek Mallick**
 - **Arav Pandey**
 - **Srinivasan Raghavan**
 - **Nidhi Mallikarjun**
 - **Ragul Narayanan Magesh**
-

1. Introduction

- EchoAI is an intelligent review response platform that helps businesses automatically monitor, analyze, and respond to customer reviews. The name "Echo" represents the system's ability to intelligently reflect back to customers, creating meaningful dialogue from their feedback. Using natural language processing and machine learning, EchoAI analyzes sentiment, generates contextually appropriate responses, and provides actionable business insights. This project demonstrates comprehensive MLOps practices including CI/CD pipelines, model versioning, monitoring, and automated retraining.

2. Dataset Information

1. **Dataset Introduction:** We utilize a combination of publicly available review datasets and real Google Reviews data to train sentiment analysis and response generation models.

2. Data Card:

- **Size:** 100,000 reviews (training), 20,000 (validation), 20,000 (test)
- **Format:** JSON/CSV
- **Language:** English (Phase 1), Spanish/French (Phase 2)
- **Features:**
 - review_text (string, 10-500 words)
 - rating (integer, 1-5 stars)
 - timestamp (datetime)
 - business_category (categorical, 20 categories)
 - review_verified (boolean)
 - helpful_votes (integer)
 - owner_response (string, when available)
 -
- **Label Distribution:** 20% (1-2 stars), 20% (3 stars), 60% (4-5 stars)

3. Data Sources:

- Google Business Profile API (OAuth authenticated)
- Yelp Open Dataset (174K businesses, 8M reviews)
- Amazon Customer Reviews (public dataset)
- Synthetic augmentation using GPT-3.5 for edge cases

4. Data Rights and Privacy:

- OAuth 2.0 consent flow for business owners
- PII detection and anonymization pipeline
- GDPR/CCPA compliant data handling

- 12-month retention policy
- Encrypted storage (AES-256)
- No reviewer personal data retained

3. Data Planning and Splits

- Ingestion:
 - Real-time: Google API webhook listeners
 - Batch: Daily pulls at 2 AM EST
- Preprocessing:
 - Text normalization (lowercase, remove URLs)
 - Spell correction using SymSpell
 - Emoji to text conversion
 - Language detection (langdetect)
 - Duplicate detection (MinHash)
- Feature Engineering:
 - TF-IDF vectorization
 - Sentiment lexicon features
 - Review length and complexity metrics
 - Temporal features (day/time patterns)
- Splitting:
 - Temporal split to prevent data leakage
 - Stratified by business category and rating
 - Train: 70% (< 6 months old)
 - Validation: 15% (6-8 months old)
 - Test: 15% (most recent 2 months)

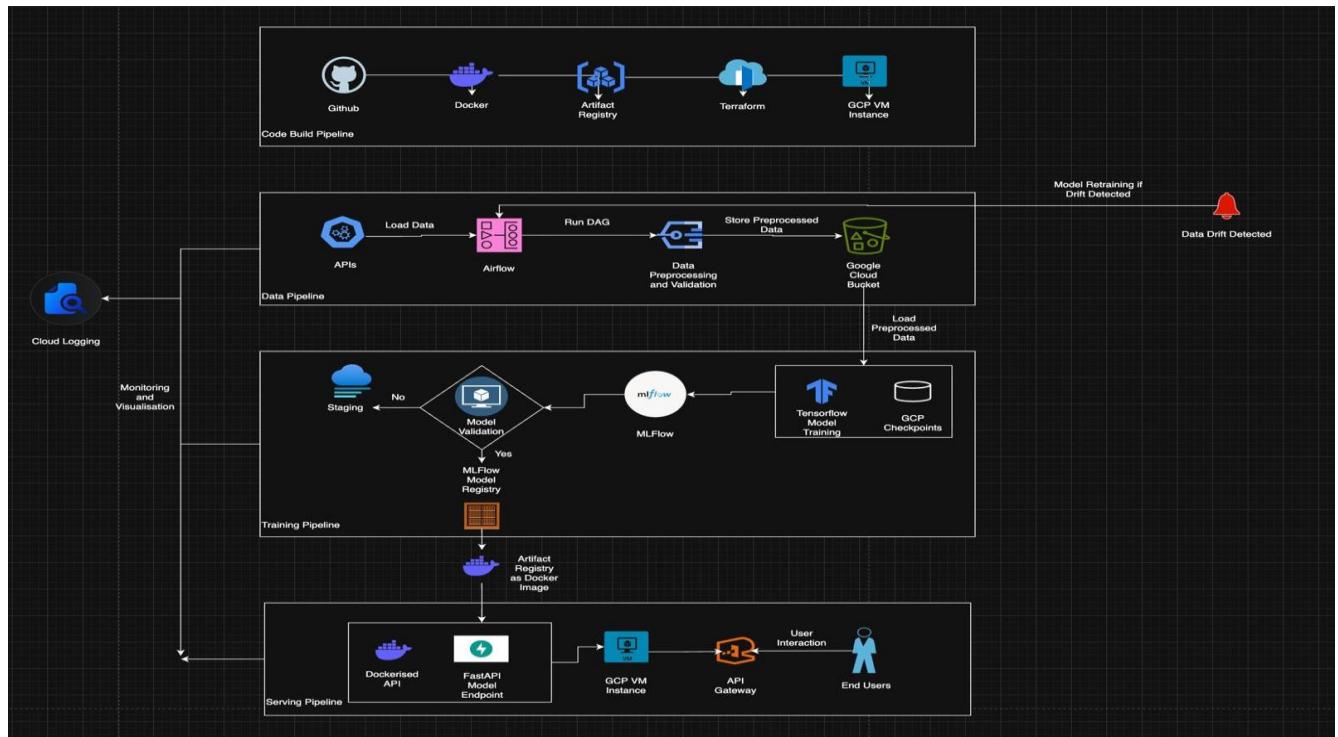
4. GitHub Repository

- **Repository:** <https://github.com/team/echo-ai>

5. Project Scope

- **5.1 Problems**
 - **Time Drain:** Businesses spend 15+ hours/week on review management
 - **Poor Response Rate:** Only 30% of reviews receive responses
 - **Delayed Responses:** Average 72-hour response time hurts SEO
 - **Generic Templates:** Copy-paste responses damage authenticity
 - **Missed Insights:** No systematic analysis of feedback patterns
- **5.2 Current Solutions**
 - Native platform dashboards (Google My Business)
 - Email notifications (often ignored/overwhelming)
 - Template libraries (impersonal)
 - Virtual assistants (\$2000+/month)
 - Review management agencies (expensive, slow)
- **5.3 Proposed Solutions**
 - **Automated Monitoring:** Real-time review aggregation via APIs
 - **Smart Prioritization:** ML-based urgency scoring
 - **Intelligent Response Generation:** Context-aware AI responses
 - **Sentiment Analysis:** Multi-aspect emotional analysis
 - **Predictive Insights:** Trend forecasting and anomaly detection
 - **One-Click Deployment:** Direct posting to platforms

6. Current Approach Flow Chart and Bottleneck Detection



7. Metrics, Objectives, and Business Goals

ML Metrics:

- **Sentiment Classification:** F1 Score > 0.85
- **Response Quality:** BLEU Score > 0.70

- **Urgency Detection:** Precision > 0.90
- **Model Latency:** < 500ms per review

Business Metrics:

- **Response Rate:** Increase to 95%
- **Response Time:** Reduce to <1 hour
- **Customer Satisfaction:** +20% improvement
- **Operational Efficiency:** 80%-time reduction
- **Cost Savings:** \$1500/month per business

Objectives:

- Deploy production system within 14 weeks
- Onboard 10 pilot businesses
- Process 10,000+ reviews in pilot phase
- Achieve 85% approval rate on generated responses

8. Failure Analysis

• Risk Matrix:

Risk	Probability	Impact	Mitigation
Inappropriate responses	Medium	High	Content filters, human review queue
API rate limiting	High	Medium	Caching, exponential backoff
Model degradation	Medium	Medium	Drift detection, auto-retraining
Data breach	Low	High	Encryption, access controls, auditing
System downtime	Low	High	Multi-region deployment, auto-scaling

Failure Handling:

- Fallback to template responses if model fails
- Manual review queue for low-confidence predictions
- Circuit breaker pattern for external APIs
- Rollback capability for model updates

9. Deployment Infrastructure: Infrastructure Stack:

Compute:

- AWS EC2 (t3.medium for API)
- AWS Lambda (for async processing)

Storage:

- PostgreSQL (RDS): Metadata
- S3: Model artifacts, logs
- Redis: Caching layer

ML Platform:

- MLflow: Model registry
- SageMaker: Model training
- FastAPI: Model serving

Monitoring:

- CloudWatch: Infrastructure
- Prometheus: Metrics
- Grafana: Dashboards

CI/CD:

- GitHub Actions: Testing
- Docker: Containerization
- Kubernetes: Orchestration

10. Monitoring Plan

Three-Layer Monitoring:

Layer 1 - System Health:

- API latency (p50, p95, p99)
- Request throughput
- Error rates
- Resource utilization

Layer 2 - Model Performance:

- Daily accuracy assessment
- Feature drift detection
- Prediction confidence distribution
- Response quality sampling

Layer 3 - Business Impact:

- Reviews processed per hour
- Response acceptance rate
- Time saved per business
- Customer satisfaction change

11. Success and Acceptance Criteria

• Technical Success:

- 99.5% uptime over 30 days
- Process 1000 reviews/minute
- Sub-second response generation
- Automated deployment pipeline

• Business Success:

- 10+ active pilot customers
- 85% generated response approval
- 90% reduction in response time
- Positive ROI within 6 months

• Acceptance Tests:

- Load testing (10x expected volume)
- Security penetration testing
- GDPR compliance audit
- User acceptance testing with 5 businesses

12. Timeline Planning

Week 1-2: Foundation

- Setup GitHub, Docker, databases
- Google API integration
- Basic data pipeline

Week 3-4: ML Models

- Train sentiment analysis model
- MLflow setup
- Model evaluation

Week 5-6: Response Generation

- Fine-tune LLM for responses
- Quality checks
- Integration with sentiment model

Week 7: Backend API

- FastAPI development
- Authentication
- Redis caching

Week 8: Frontend & Integration

- React dashboard
- Connect all components
- Testing

Week 9: Deployment

- Docker/Kubernetes setup
- Deploy to cloud
- Documentation & monitoring

13. Additional Information

Competitive Advantage:

- First to market with LLM-powered responses
- Industry-specific model fine-tuning
- Real-time processing capabilities

Ethical Guidelines:

- Transparent AI disclosure to customers
- Human-in-the-loop for sensitive topics
- Bias auditing quarterly
- No manipulation of review content

Scaling Strategy:

- Horizontal scaling via Kubernetes
- Multi-tenant architecture
- Regional deployment for latency