STEMBlock.ai - Executive Summary and Project Overview

Project Information

Project Name: STEMBlock.ai - AI-Based Automated STEM Evaluation System **Domain**: https://www.stemblock.ai **API Endpoint**: https://api.stemblock.ai **Version**: 1.0 **Date**: October 18, 2025

---

1. Executive Summary

STEMBlock.ai is a comprehensive AI-powered platform designed to revolutionize STEM education assessment, specifically focused on VEX robotics (IQ and V5) programs. The system leverages Google Vertex AI to provide automated, intelligent evaluation of student work including robot designs, programming code, and engineering notebooks, while maintaining a collaborative learning environment through integrated forum and reporting features.

1.1 Core Value Proposition

- **Automated Evaluation**: AI-powered assessment reduces coach workload by 50% while providing immediate, detailed feedback to students
- **Personalized Learning**: Adaptive assessments and tailored feedback help each student improve at their own pace
- **Data-Driven Insights**: Comprehensive analytics and benchmarking help coaches, parents, and students understand progress and growth areas
- **VEX Robotics Specialization**: Purpose-built for VEX competitions with deep understanding of rules, best practices, and evaluation rubrics

1.2 Key Differentiators

1. **AI Model Benchmarking IP**: Proprietary evaluation dataset and benchmarking framework for assessing language models on STEM education tasks
2. **Multi-Modal AI Analysis**: Evaluates images (robot designs), code (Python/C++/Blocks), and documents (engineering notebooks)
3. **Role-Based Experience**: Tailored interfaces for students, coaches, parents, and administrators
4. **Continuous Improvement**: AI learns from expert feedback and production data to improve evaluation quality over time

---

2. Team Structure and Timeline

2.1 Team Composition

Total Team: 4 people

Role	Count	Responsibilities
Developers	2	Full-stack development (Frontend + Backend + AI integration)
UX Designer	1	UI/UX design, user research, wireframes, prototypes
Intern	1	AI model evaluation, dataset creation, testing, documentation

2.2 Timeline

Total Duration: 6 months

Phase	Duration	Milestone
MVP	3 months	Core features: Authentication, Submission & Evaluation, Basic Reporting
Full v1.0 Release	Additional 3 months	All features: Forum, Analytics, Benchmarking, Polish

MVP Scope (Month 1-3): - User authentication and management - File upload and submission - AI-powered evaluation (basic) - Individual student reports - Basic coach analytics - Essential RBAC

Full Release Scope (Month 4-6): - Forum and Q&A system - Advanced reporting and benchmarking - Assessment system - AI model evaluation benchmark - Performance optimization - Security hardening - Production launch

---

3. Technical Architecture

3.1 Technology Stack

Frontend: - Next.js 14+ (React 18, TypeScript) - Tailwind CSS + shadcn/ui - React Query + Zustand

Middleware: - Go 1.21+ microservices - Gin or Echo framework - Clean architecture pattern

Database: - PostgreSQL 15+ (primary) - Redis 7+ (cache, sessions)

AI Platform: - Google Vertex AI - Gemini Pro (text analysis) - Gemini Pro Vision (image analysis) - Prompt caching for cost optimization

Cloud Infrastructure: - Google Cloud Platform (GCP) - Cloud Run (serverless containers) - Cloud SQL (managed PostgreSQL) - Memorystore (managed Redis) - Cloud Storage (file storage) - Cloud CDN (static assets)

3.2 Microservices Architecture

1. **API Gateway**: Request routing, authentication, rate limiting
2. **Auth Service**: User authentication, session management, OAuth
3. **User Service**: User profiles, RBAC, class/team management
4. **Assessment Service**: Dynamic question generation, test administration
5. **Evaluation Service**: File processing, AI evaluation orchestration
6. **Report Service**: Analytics, benchmarking, PDF generation
7. **Forum Service**: Discussion board, Q&A, voting
8. **Notification Service**: Email, in-app, WebSocket notifications

3.3 Key Design Decisions

- **Serverless Architecture**: Cloud Run for auto-scaling and cost efficiency
- **Microservices**: Independent scaling and deployment of services
- **JWT Authentication**: Stateless authentication with refresh token rotation
- **RBAC**: Four-tier role system (Student, Parent, Coach, Admin)
- **Vertex AI**: Native GCP integration, prompt caching, latest models
- **Infrastructure as Code**: Terraform for reproducible deployments

---

4. Core Features

4.1 Assessment System

- Dynamic question generation with adaptive difficulty
- Support for multiple choice, short answer, code, and image submissions
- Automated AI grading for objective and subjective questions
- Real-time feedback and results

4.2 Evaluation Engine

- **Multi-modal Analysis**:
  - Images: Robot design evaluation, component identification
  - Code: Static analysis + AI review with optimization suggestions
  - Documents: Engineering notebook assessment against VEX rubrics
- **AI Pipeline**: Async processing with background workers
- **Feedback Generation**: Constructive, specific, age-appropriate feedback

4.3 Reporting and Analytics

- **Individual Reports**: STEM capability scores, progress tracking, benchmarking
- **Coach Analytics**: Class performance, at-risk student identification, engagement metrics
- **Parent Reports**: Simplified, actionable insights on student progress
- **Benchmarking**: Compare to class, grade, organization averages

4.4 Forum and Collaboration

- Reddit-style discussion board with Q&A functionality
- Threaded comments (up to 5 levels)
- Upvote/downvote system
- Tag-based organization
- Content moderation tools

4.5 AI Model Evaluation (IP)

- Proprietary benchmark dataset (1,000+ annotated samples)
- Evaluation framework for assessing LLM performance on STEM tasks
- Monthly model performance tracking
- Quarterly comprehensive benchmark reports
- Similar to Cybench and CyberGym but for STEM/VEX robotics

---

5. Security and Compliance

5.1 Security Measures

Authentication: - bcrypt password hashing (cost factor 12) - JWT with RS256 signing - Dual-token strategy (access + refresh) - Session management in Redis - OAuth 2.0 (Google) with PKCE

Authorization: - Role-Based Access Control (RBAC) - Four roles: Student, Parent, Coach, Admin - Granular permissions per resource - Parent-student linking requires approval

Data Protection: - TLS 1.3 encryption in transit - AES-256 encryption at rest - PII field-level encryption - Signed URLs for file access - Secrets in Google Secret Manager

Application Security: - Input validation and sanitization - SQL injection prevention (parameterized queries) - XSS protection (React escaping + CSP) - CSRF protection - Rate limiting (per user and per IP) - Web Application Firewall (Cloud Armor)

5.2 Compliance

FERPA (Family Educational Rights and Privacy Act): - Protect student educational records - Audit trails for data access - Parent/student data access portal - Data encryption and access controls

COPPA (Children's Online Privacy Protection Act): - Age verification for users under 13 - Parental consent workflow - Limited data collection for children - Parent data access and deletion rights

GDPR-Ready (General Data Protection Regulation): - User consent management - Right to access, erasure, portability - Data breach notification procedures - Privacy by design

---

6. Deployment and Infrastructure

6.1 Environments

- **Development**: Local Docker Compose, seeded data
- **Staging**: GCP Cloud Run, mirrors production, auto-deploy from `develop` branch
- **Production**: GCP Cloud Run, manual approval for deployment, stemblock.ai domain

6.2 CI/CD Pipeline

Tool: GitHub Actions

Workflows: 1. **CI Pipeline**: Lint, test, security scan, build (on every PR) 2. **Deploy Staging**: Auto-deploy to staging on merge to `develop` 3. **Deploy Production**: Manual approval required for merge to `main`

Deployment Strategy: - Blue-green deployment (Cloud Run native) - Zero-downtime deployments - Automated rollback on health check failure

6.3 Monitoring and Alerting

Monitoring: - Google Cloud Monitoring (system metrics) - Cloud Logging (centralized logs) - Cloud Trace (distributed tracing) - Uptime checks - Custom business metrics

Alerting: - PagerDuty for critical alerts (service down, high error rate) - Slack for high-priority alerts (performance degradation) - Email for medium-priority alerts - 24/7 on-call rotation

6.4 Estimated Costs

Monthly Infrastructure (1,000 users): ~$2,000 - Cloud Run: $550 - Cloud SQL: $350 - Redis: $200 - Storage: $100 - Load Balancer + CDN: $130 - Vertex

AI: $500 - SendGrid: $90

Scaling: - 5,000 users: ~$5,000/month - 10,000 users: ~$10,000/month

---

7. AI Model Evaluation Benchmark (Proprietary IP)

7.1 Purpose

Create a comprehensive benchmark for evaluating language models on STEM education and VEX robotics tasks, establishing STEMBlock.ai as a thought leader and creating valuable intellectual property.

7.2 Dataset

- **Size**: 1,000+ annotated samples by Month 6
- **Categories**:
  - Programming submissions (500 samples)
  - Robot design images (300 samples)
  - Engineering notebooks (200 samples)
  - Short answer questions (400 samples)
  - Edge cases (100 samples)
- **Annotation**: Expert VEX coaches provide ground truth evaluations

7.3 Models to Evaluate

- Google Gemini 1.5 Pro/Flash
- OpenAI GPT-4o/GPT-4o-mini
- Anthropic Claude 3.5 Sonnet/Haiku
- Ongoing evaluation of new models

7.4 Metrics

- **Accuracy**: % within 10 points of human expert
- **F1 Score**: Precision and recall for issue detection
- **Response Time**: Latency per evaluation
- **Cost Efficiency**: Cost per evaluation
- **Feedback Quality**: Evaluated by meta-LLM judge (1-5 scale)

7.5 Deliverables

- **Public Benchmark Report**: Methodology, sample results, model leaderboard
- **Internal Dashboard**: Real-time model performance tracking
- **Research Publications**: Conference papers, blog posts
- **Proprietary Assets**: Full dataset, detailed prompts, evaluation pipeline

7.6 Team Responsibility

- **Intern (50% time)**: Dataset creation, annotation, experiment execution, report compilation
- **Developer 1 (20% time)**: Evaluation pipeline, API integration, dashboard
- **Developer 2 (10% time)**: Data collection tools, annotation interface
- **UX Designer (5% time)**: Visualization and report design

---

8. Implementation Roadmap (6 Months)

Phase 0: Foundation (Weeks 1-4)

- Infrastructure setup (GCP, Terraform)
- Repository structure and CI/CD
- Development environment
- Initial AI prompt templates

Phase 1: Core Platform (Weeks 5-12) - MVP TARGET

- Authentication and user management
- Class and team management
- File upload system
- Basic AI evaluation (code + images)
- Individual student reports
- Coach analytics dashboard
- **Deliverable**: Functional MVP ready for pilot testing

Phase 2: Assessment System (Weeks 13-16)

- Question management
- Assessment delivery
- Adaptive difficulty
- Automated scoring
- AI evaluation for short answers

Phase 3: Advanced Features (Weeks 17-20)

- Engineering notebook analysis
- Advanced reporting and benchmarking
- Forum and Q&A system
- Notification system

Phase 4: AI Model Benchmark (Weeks 1-24, parallel)

- Dataset creation and annotation (ongoing)

- Evaluation pipeline development
- Monthly model evaluations
- Quarterly comprehensive benchmark reports
- **Deliverable**: Published benchmark report by Month 6

Phase 5: Polish and Launch (Weeks 21-24)

- Performance optimization
- Security hardening
- User testing and refinement
- Documentation
- Production launch
- **Deliverable**: Full v1.0 production release

---

9. Success Criteria

9.1 MVP Success (Month 3)

☐ 100+ students onboarded in pilot program
☐ 80%+ user satisfaction score
☐ < 5 minutes average AI evaluation time
☐ All core features functional
☐ Security audit passed

9.2 Full Release Success (Month 6)

☐ 500+ active users
☐ 80%+ weekly active user rate
☐ 90%+ user satisfaction (NPS > 50)
☐ Coaches report 50% time savings
☐ AI evaluation accuracy > 85%
☐ System uptime > 99%
☐ AI benchmark report published
☐ Zero critical security vulnerabilities

9.3 Business Goals (Year 1)

☐ 1,000+ active users
☐ 5+ schools/organizations onboarded
☐ 3+ media mentions or articles
☐ Break-even on infrastructure costs
☐ Partnership inquiries from AI vendors

---

10. Risk Management

10.1 Key Risks

Risk	Mitigation
Small team, aggressive timeline	Ruthless prioritization, leverage managed services, no scope creep
AI evaluation quality	Multi-model evaluation, human review sampling, continuous monitoring
Vertex AI costs	Prompt caching, budget alerts, cost-effective model selection
User adoption	Early pilot program, user testing, coach involvement, training materials
Security vulnerabilities	Security-first development, regular audits, penetration testing
Scope creep	Strict MVP definition, defer Phase 2 features, change control process

10.2 Contingency Plans

- **If timeline slips**: Reduce MVP scope further, focus on core evaluation only
- **If primary AI model underperforms**: Switch to best alternative (GPT-4o or Claude)
- **If costs exceed budget**: Use more cost-effective models, optimize prompts, reduce features
- **If pilot program delayed**: Use synthetic data and team testing

---

11. Document Structure

This comprehensive planning package consists of 8 detailed documents:

1. **Executive Summary** (this document): High-level overview and key decisions
2. **Requirements Document**: Detailed functional and non-functional requirements
3. **System Architecture**: Technical architecture, data flow, component design
4. **Technical Implementation Plan**: Phase-by-phase development plan with tasks
5. **API Specifications**: Complete API documentation with endpoints and examples
6. **Database Schema**: Comprehensive database design with tables, indexes, and relationships

7. **Security and RBAC Design**: Authentication, authorization, compliance, and security measures
8. **Deployment and Infrastructure**: Cloud infrastructure, CI/CD, monitoring, and cost management
9. **AI Model Evaluation Plan**: Benchmarking framework, dataset, and evaluation methodology

---

12. Next Steps

Immediate Actions (Week 1)

1. **Team Kickoff Meeting**: Review all documents, assign responsibilities, set up communication channels
2. **GCP Setup**: Create project, enable APIs, set up billing alerts, configure IAM
3. **Repository Setup**: Create GitHub organization, repositories, branch protection
4. **Development Environment**: Docker Compose setup, database schemas, seed data
5. **Design Kickoff**: UX designer begins wireframes for core user flows

Week 2-4 Priorities

1. **Infrastructure**: Complete Terraform configurations, deploy to dev environment
2. **Frontend Scaffolding**: Next.js project setup, UI component library, routing structure
3. **Backend Services**: Implement auth service, API gateway, basic database operations
4. **AI Integration**: Set up Vertex AI, create initial prompt templates, test API calls
5. **CI/CD**: Configure GitHub Actions workflows, deployment pipelines

Month 2 Focus

1. **MVP Core Features**: Authentication, user management, file upload, basic evaluation
2. **AI Evaluation Pipeline**: Implement evaluation service, background workers, prompt optimization
3. **UX Testing**: Conduct early user testing with wireframes and prototypes
4. **Dataset Creation**: Begin creating synthetic samples for AI model evaluation

Month 3 - MVP Launch

1. **Feature Completion**: All MVP features implemented and tested

2. **Pilot Program**: Onboard first 100 students and coaches
3. **Monitoring**: Set up comprehensive monitoring and alerting
4. **Security**: Complete security audit, penetration test
5. **Documentation**: User guides, API docs, operational runbooks

---

13. Conclusion

STEMBlock.ai represents a unique opportunity to leverage cutting-edge AI technology to transform STEM education assessment and student learning outcomes. With a focused MVP in 3 months and full release in 6 months, the project is ambitious but achievable with the right prioritization and execution.

Key Success Factors: 1. **Ruthless Prioritization**: Focus on core value proposition, defer nice-to-haves 2. **Leverage Managed Services**: Use GCP managed services to reduce operational burden 3. **AI-First Approach**: Invest in AI evaluation quality and benchmarking IP 4. **User-Centric Design**: Continuous feedback from coaches and students 5. **Security and Compliance**: Build trust through robust security and privacy measures 6. **Iterative Development**: Ship MVP, gather feedback, iterate quickly

With a talented team, clear vision, and comprehensive planning, STEMBlock.ai is positioned to become the leading AI-powered STEM evaluation platform for VEX robotics education.

Project Start Date: Week 1, 2025 **MVP Target**: Month 3 **Full Release Target**: Month 6 **Status**: Planning Complete, Ready for Development

Prepared by: AI Planning Assistant **Date**: October 18, 2025 **Version**: 1.0

For questions or clarifications, refer to the detailed planning documents.