

CODE LENS v2 - Enterprise Application Intelligence Platform

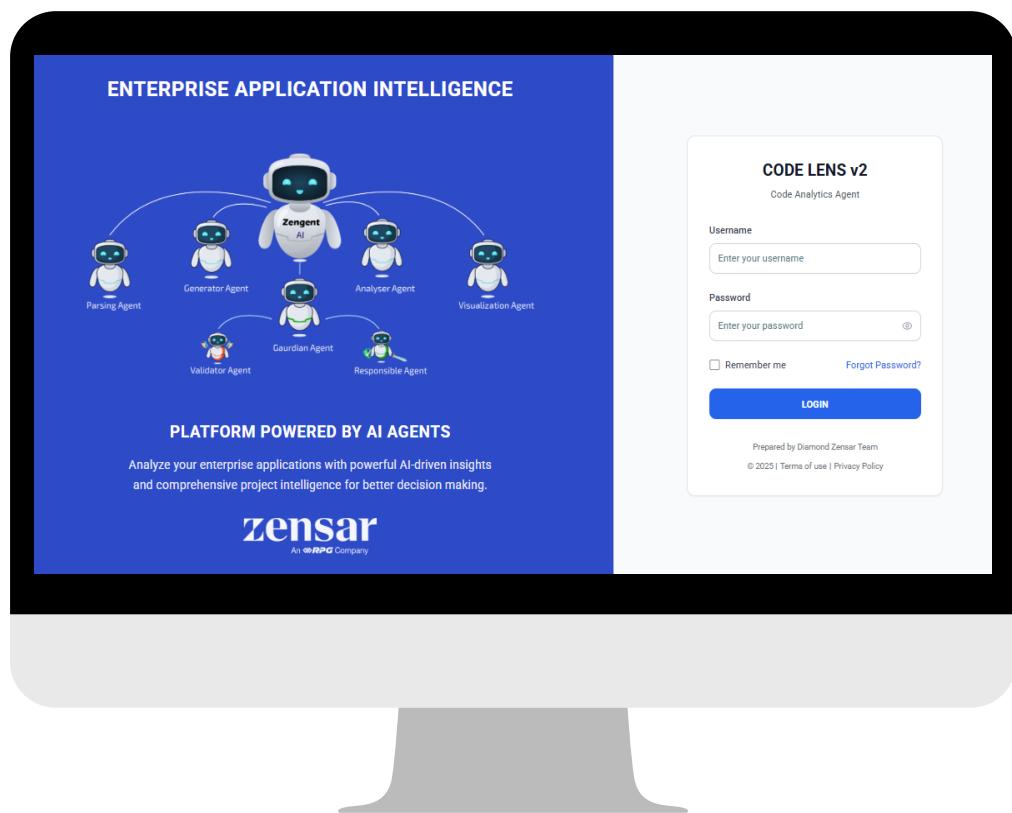
Platform Tagline: Code Analytics Agent

Team Member : Kaushik Saha, Sameer Kumar Sharma, Sachin Nilawar, Satya Yenamandra, Piyush Gupta, Ullas Krishnan, Hrushikesh Nalawade

Project : AMEX Diamond

Project Description

CODE LENS v2 is a comprehensive enterprise application intelligence platform that analyses multi-language project architectures from ZIP uploads and GitHub repositories. It supports Java, Python, PySpark, Mainframe (COBOL/JCL), C#, and Kotlin codebases, extracting and parsing source files to identify architectural patterns, dependencies, and relationships. The platform provides interactive visual diagrams and AI-powered analysis using various LLMs (OpenAI GPT-4o, Ollama local models), delivering intelligent insights and recommendations.



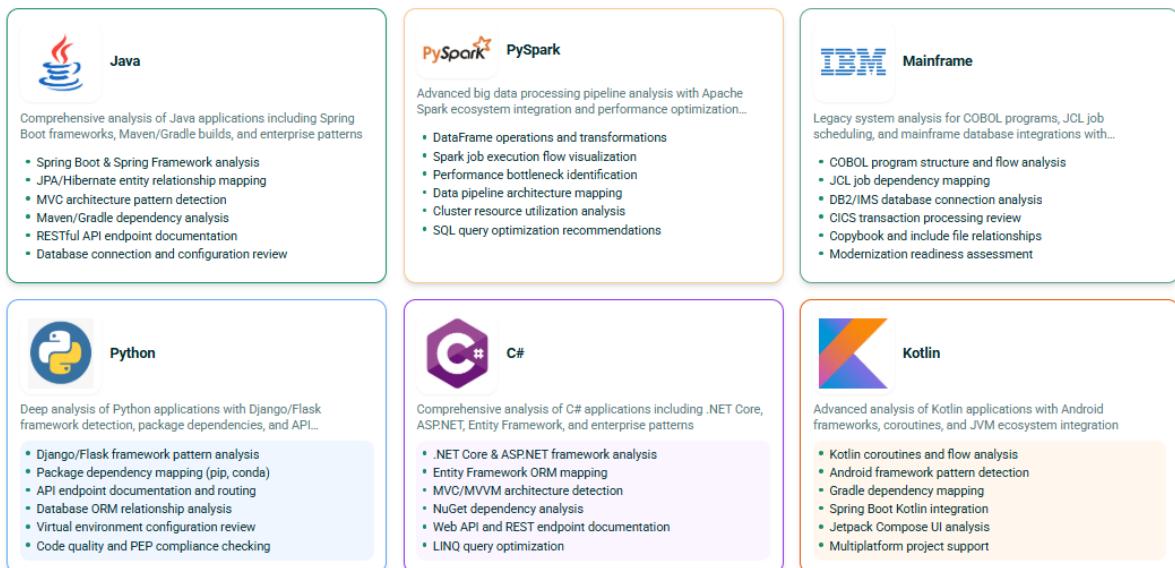
With enterprise features including Swagger API documentation, demographic field scanning with compliance reporting, integration pattern detection (30+ patterns), POD→POA (AS-IS→TO-BE) migration planning, **Data Image and Quality Measures** analysis, professional PDF/DOC/HTML report generation with corporate branding, and upcoming **Agentic AI capabilities** using LangChain/LangGraph, CODE LENS

v2 provides comprehensive, AI-driven insights into complex enterprise applications to aid development, refactoring, and strategic planning. The vision is to accelerate code analysis by 85%, ensure data privacy compliance, and empower teams with actionable intelligence—all 100% offline for maximum security.

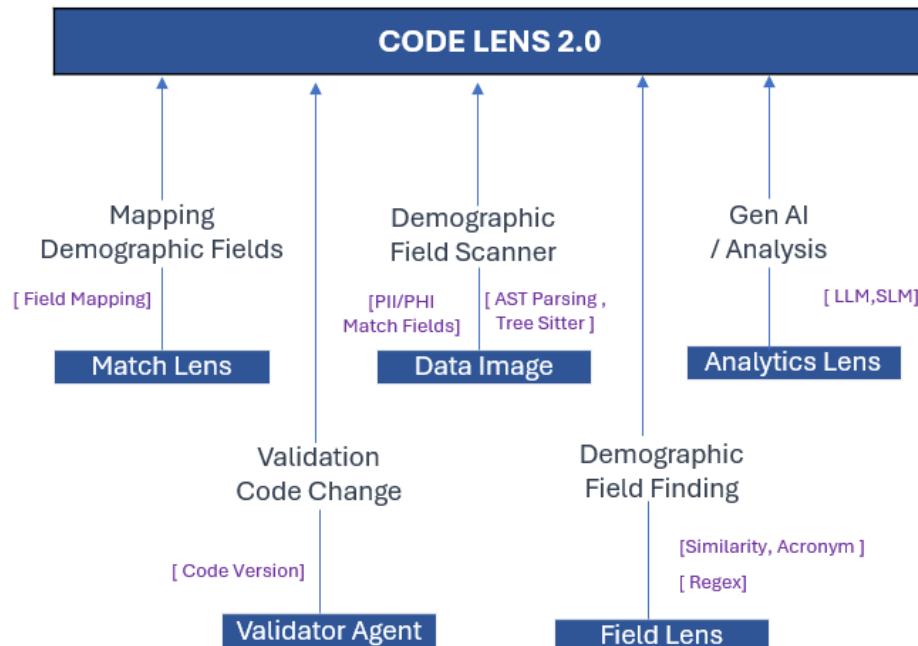
Enterprise Application Code Scan AI ML Utilities

Advanced multi-language code analysis utilities supporting Java, Python, PySpark, Mainframe, C#, and Kotlin. Leverage AI/ML capabilities for comprehensive application intelligence, architectural insights, and automated code scanning across enterprise frameworks and technologies.

[Java & Spring Boot](#) [Python & Django](#) [PySpark & Big Data](#) [Mainframe & COBOL](#) [C# & .NET](#) [Kotlin & Android](#) [AI/ML Analysis](#)



Enterprise Application Intelligence



Key Benefits

ASSIST

- Helps developers understand legacy codebases, plan migration strategies, and ensure data privacy compliance
- Provides AI-powered recommendations for modernization and refactoring
- Identifies integration patterns and technical debt for informed decision-making

AUGMENT

- Enhances human analysis with multi-LLM AI insights (GPT-4o, Claude, Gemini, local models)
- Interactive diagrams and visual architecture representations
- Intelligent pattern recognition across 30+ integration points
- Demographic data analysis with compliance risk assessment

AUTOMATE

- Automatically scans 30+ demographic field patterns and integration patterns
- Generates architecture diagrams, API documentation, and migration roadmaps
- Creates professional PDF reports with corporate branding
- Extracts dependencies, relationships, and architectural insights—eliminating weeks of manual work

Value Proposition

- **70-80% Time Savings:** Reduce manual code analysis effort from 40 hours to 8 hours for 10K+ line projects
- **95%+ Compliance Detection:** Comprehensive PII/PHI field detection for GDPR, HIPAA, PCI-DSS compliance
- **100% Offline Capable:** Complete data security with local LLM models and privacy-first architecture
- **Multi-Language Support:** Single platform for Java, Python, PySpark, Mainframe, C#, Kotlin analysis
- **40-60% Cost Reduction:** Lower migration costs through AI-powered phased planning and risk identification
- **6-12 Month ROI:** Break-even timeline with \$200K-500K annual savings post-migration

Quantitative Benefits & ROI Metrics

Metric	Value	Impact
Time Savings	70-80%	Manual review of 10K+ line codebase: 40 hours → 8 hours
Cost Reduction	40-60%	Typical \$500K migration project reduced to \$200-300K
Compliance Detection	95%+ accuracy	Identifies 95%+ of PII/PHI fields vs 60-70% manual review
Documentation Speed	90% faster	Architecture documentation: 20 hours → 2 hours
Code Quality Improvement	30-50%	Technical debt reduction through AI recommendations
Migration ROI	6-12 months	Break-even timeline with annual savings of \$200K-500K
Risk Mitigation	80%+	Prevents data loss/security incidents through pre-migration scanning
Developer Productivity	2-3x increase	New developer ramp-up: 2-3 weeks → 3-5 days

Qualitative Benefits

Strategic Decision Making

- Data-driven modernization planning with AI-powered insights
- Clear visibility into technical debt and architectural risks
- Informed technology stack selection based on actual codebase analysis
- Compliance confidence with automated PII/PHI detection

Risk Reduction

- Early identification of security vulnerabilities (CWE patterns)

Zensar Technologies

OCT 2025 AES – Diamond Project Team

Mentor : kaushik.saha@zensar.com
 : sameerkumar.sharma@zensar.com

- Compliance violation detection before production deployment
- Dependency risk assessment and cyclic dependency breaking
- Migration risk mitigation with phased roadmap planning

Team Collaboration

- Shared understanding of complex legacy systems
- Visual architecture diagrams for stakeholder communication
- Comprehensive documentation reducing knowledge silos
- Onboarding acceleration for new team members

Innovation Enablement

- Modern architecture recommendations (microservices, cloud-native, serverless)
- Technology upgrade paths with clear migration strategies
- Best practice identification and anti-pattern detection
- AI-driven code quality improvements

Zengent - Gen AI Agents for Diamond Project

Specialized AI agents for advanced analysis and processing

Gen AI - Agents for Diamond Project

Three specialized AI agents for enterprise development: Knowledge Agent for document intelligence and Q&A with Confluence integration, Code Conversion Agent for multi-language code transformation, and Validator & Responsible AI Agent for security compliance, privacy validation, ethics assessment, and bias detection across your applications.

Knowledge Intelligence Code Conversion Security & Compliance Responsible AI



Knowledge Agent

Document intelligence and Q&A system with Confluence integration, PDF processing, and intelligent knowledge extraction

- Confluence web scraping integration
- IBM Doclinc PDF processing
- LangChain document processing
- Redis caching for performance
- Intelligent Q&A chat interface
- Multi-source knowledge aggregation



Code Conversion Agent

Multi-language code conversion agent for transforming source code between programming languages while preserving logic

- Multi-language code translation
- Framework migration support
- AST-based syntax conversion
- Code quality preservation
- Bulk project processing
- Language ecosystem mapping



Validator & Responsible AI Agent

Comprehensive validation covering security vulnerabilities, privacy compliance, ethics, bias detection, and AI governance

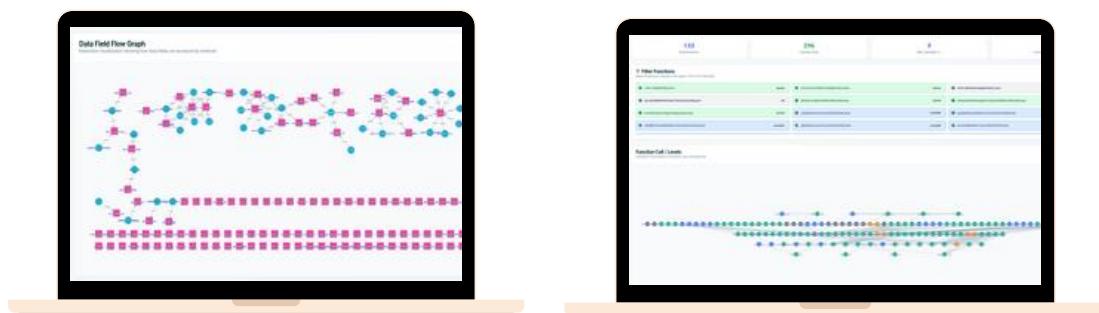
- Security vulnerability detection and assessment
- Privacy compliance and data protection validation (GDPR,...
- AI ethics and bias detection in algorithms
- Fairness assessment across demographic groups
- Code quality metrics and best practices review
- Algorithmic transparency and explainability
- Responsible AI deployment and governance validation
- Performance bottleneck identification
- Security risk scoring and remediation guidance

Key Features

1. Multi-Language Code Analysis

- Support for Java (Spring Boot, JPA, MVC), Python (Django/Flask), PySpark (DataFrame, job flow), Mainframe (COBOL, JCL), C#, Kotlin
- Automatic ZIP extraction and validation (up to 50MB)
- GitHub integration for direct repository import
- Handles large enterprise applications (1000+ files)
- Intelligent code parsing and structure analysis

2. Demographic Field Scanning & Compliance



Demographic Field Scanning & Compliance

Function Level – Call



Two Scanning Paths:

Zensar Technologies
OCT 2025 AES – Diamond Project Team

Mentor : kaushik.saha@zensar.com
: sameerkumar.sharma@zensar.com

- **Regex Scan:** 39 pre-defined patterns across 5 categories (Primary Names, Secondary Names, Government IDs, Contact Information, Financial/Health Data)
- **Excel Field Mapping:** Upload custom field lists with independent field/table name matching

Capabilities:

- Detects sensitive data: SSN, phone numbers, emails, addresses, birth dates, credit cards, medical records
- Shows exact file locations, line numbers, and code context
- Identifies which classes and functions use demographic fields
- Separates business logic classes from data-only DTOs/Entities/Models
- SQL query analysis for database field tracking
- Compliance reporting with detailed coverage metrics

3. Demographic Class Analysis

- Identifies classes containing demographic fields with business logic functions
- Function-level analysis showing how sensitive data is processed
- Smart classification excludes DTOs/Entities without functions
- Diagnostic reporting explains why files were omitted
- Detailed function signatures and descriptions

4. AI-Powered Insights & Recommendations

Online Mode:

- OpenAI GPT-4o integration for cloud-based analysis

Offline Mode:

- Local LLMs via Ollama (Code Llama, Deepseek Coder, StarCoder, Llama 3, Mistral)
- Complete data privacy—no external dependencies

Capabilities:

- Project overview and architecture analysis
- Actionable recommendations for improvement
- Quality assessment scores
- Module-level insights for individual classes
- Code quality analysis with HuggingFace CodeBERT

5. Code Lens ML (Traditional Machine Learning)

- Scikit-learn based field suggestions (NOT LLM)
- Activated on explicit user request via "Get ML Suggestions" button
- Pattern-based demographic field recommendations
- Pure Python/NumPy implementation for offline operation

6. Swagger API Documentation

- Automatic extraction from REST controllers
- Identifies endpoints, HTTP methods, request/response models
- Generates interactive API documentation
- Supports Spring Boot annotations (@RestController, @RequestMapping, @GetMapping, etc.)

7. Interactive Visual Diagrams

- UML-style class diagrams with AntV X6 and React Flow
- Dependency graphs and architectural relationships
- Language-specific representations

- Mermaid.js integration for flow diagrams
- Component, sequence, and comprehensive diagrams

8. Professional Report Generation

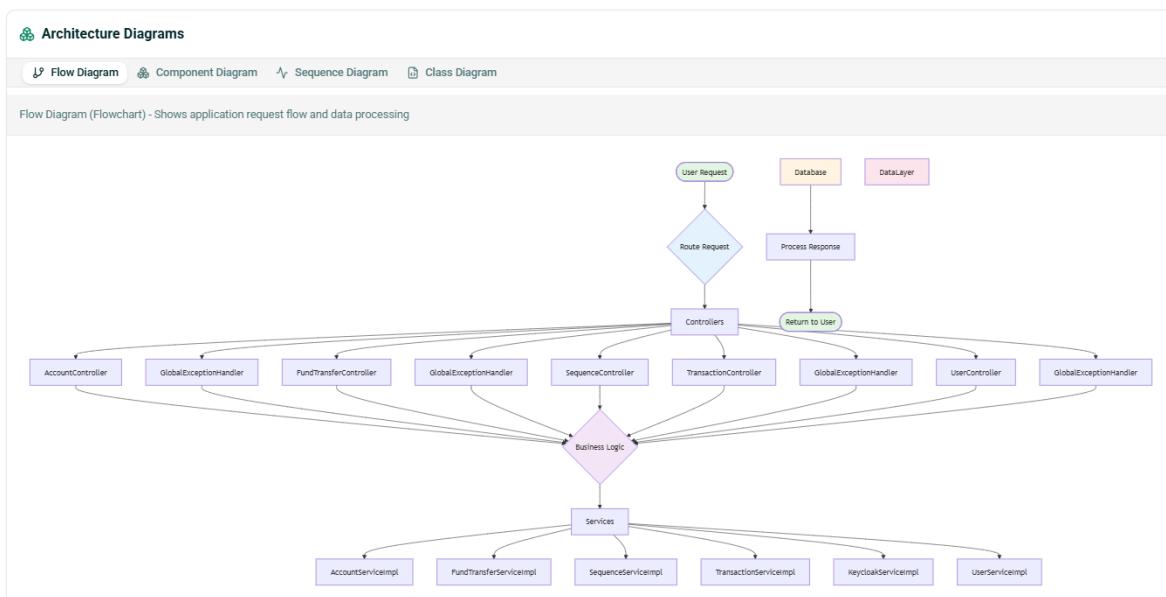
- Export formats: PDF, DOC, HTML
- Comprehensive analysis with diagrams and insights
- Corporate branding footer: "Developed by: Ullas Krishnan, Sr Solution Architect. Copyright © Project Diamond Zensar team"
- HTML preview before download
- Customizable report sections

9. Vector Database Intelligence (ZenVector Agent)

- ChromaDB for persistent vector storage
- Code similarity detection and semantic search
- Demographic data analysis with vector embeddings
- Multi-modal search capabilities
- Sentence transformer models for semantic understanding

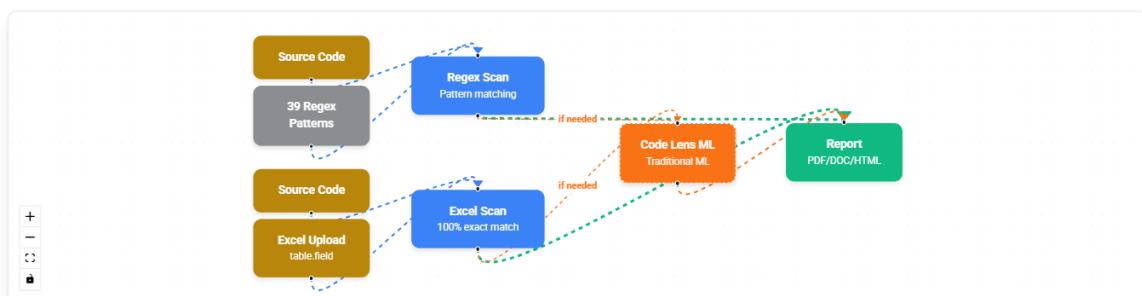
10. Knowledge Agent - Document Intelligence & Q&A

- Confluence integration for enterprise documentation
- IBM Doclq for PDF processing
- LangChain for document processing workflows
- LangGraph for AI workflow orchestration
- Langfuse for LLM observability and tracing
- Redis caching for performance optimization
- Chat interface for context-aware Q&A

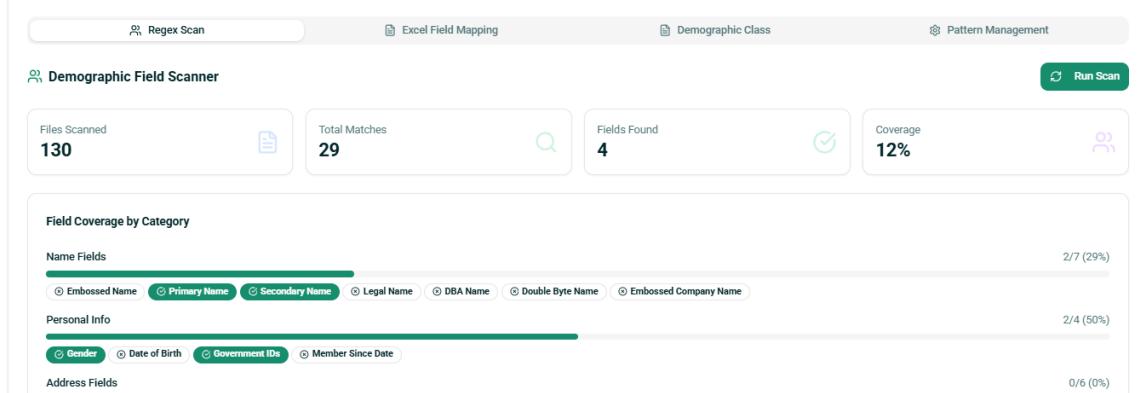


Code Lens - Demographic Data Scanning Process

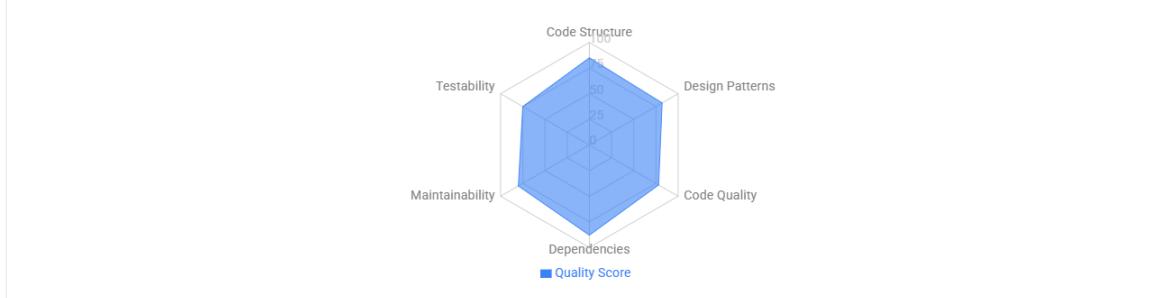
Automated PII/PHI detection with compliance reporting



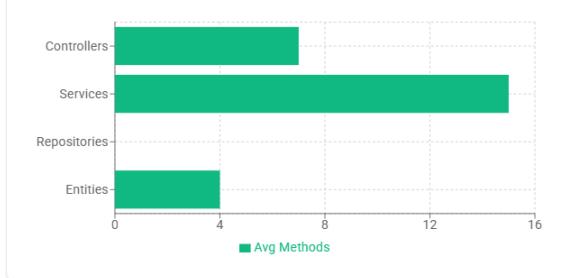
 Code Lens - Demographic Scan



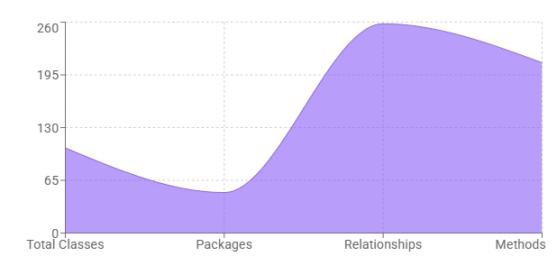
◎ Architecture Quality Metrics



<> Methods per Component Type



Package & Relationship Metrics



 Spring-Boot-Microservices-Banking-Application
Analyzed 130 Java files • Found 9 controllers, 6 services, 0 repositories

Analyzed 130 Java files • Found 9 controllers, 6 services, 0 repositories

 View Report

 Demographic Report

AI Insights

Migration Planner

9

6 Services

0
Repositories

6

Zensar Technologies
OCT 2025 AES – Diamond Project Team

Mentor : kaushik.saha@zensar.com
: sameerkumar.sharma@zensar.com

11. Integration Pattern Detection (30+ Patterns)

Comprehensive pattern detection across 5 categories:

API Integration Patterns (6 patterns):

- Synchronous Request-Response (RPC Style) detection
- Monolithic REST/SOAP endpoints identification
- Point-to-Point HTTP calls tracking
- SOAP Web Services analysis
- Service discovery gap detection
- API Gateway absence identification

Database Integration Patterns (5 patterns):

- Shared Database Pattern detection
- Integration via Database Tables
- Stored Procedure Integration tracking
- DB Vendor Lock-in (Oracle, DB2, MSSQL) identification
- Direct SQL from UI/Controllers detection

Messaging & Event-Driven Patterns (7 patterns):

- Legacy Message Queue (MQ/JMS/MSMQ) detection
- Point-to-Point Queue Pattern identification
- Fire-and-Forget Messages tracking
- Event Sourcing gap detection
- Synchronous Message Processing identification
- Dead Letter Queue (DLQ) absence detection
- Tight Queue Coupling tracking

File & Batch Integration Patterns (5 patterns):

- Scheduled Batch Jobs (Quartz, @Scheduled) detection
- SFTP/FTP File Transfer tracking
- File Polling Pattern identification
- CSV/Excel Batch Processing detection
- Large File Processing tracking

Cross-Cutting Patterns (7 patterns):

- HTTP Client Libraries detection
- Logging Frameworks identification (Log4j, Logback, SLF4J)
- Caching Solutions tracking (Ehcache, Hazelcast, Redis)
- Security Configurations analysis
- Configuration Management detection
- Monitoring & Metrics gap identification
- Transaction Management tracking

Capabilities:

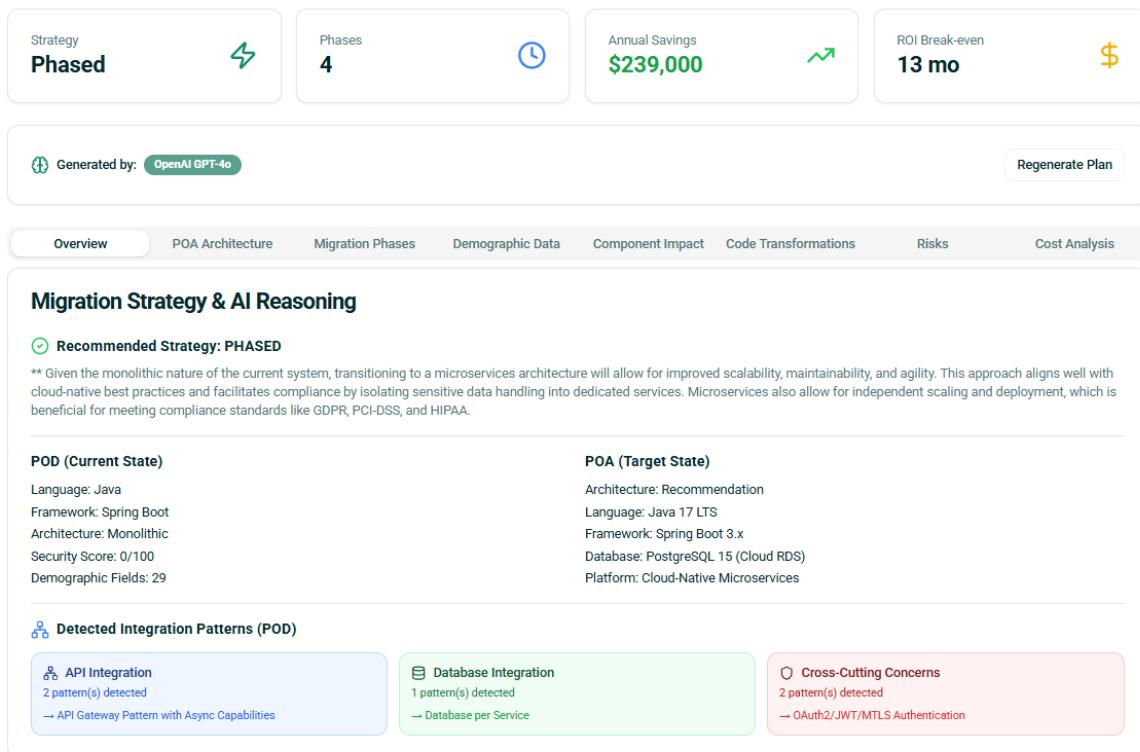
- Automated pattern scanning with regex-based code analysis
- Evidence collection with file locations and line numbers
- POA (Point of Arrival) recommendations for each detected pattern
- Modernization priority assessment (LOW, MEDIUM, HIGH, CRITICAL)
- Visual pattern detection summary showing all scanned patterns

12. POD

→ POD → POA Migration Planner

AI-powered migration path from Point of Departure to Point of Arrival

Spring-Boot-Microservices-Banking-Application



→POA Migration Planning (AI-Powered)

Intelligent migration planning from legacy (POD) to modern (POA) systems:

POD (Point of Departure) Assessment:

- Automated legacy system analysis using dependency graphs
- Demographic field scanning for compliance-critical components
- ISO-5055 quality metrics and CWE vulnerability assessment
- Integration pattern detection and technical debt identification
- Annual operating cost estimation (infrastructure, licenses, compliance risk)

POA (Point of Arrival) Recommendations:

- AI-powered architecture suggestions (microservices, cloud-native, serverless)
- Technology stack modernization (language upgrades, frameworks, databases, cloud platforms)
- Migration strategy selection with AI reasoning:
 - Strangler Fig Pattern:** Gradual service extraction
 - Phased Migration:** Multi-phase approach
 - Big Bang Migration:** Complete replacement
 - Lift & Shift:** Infrastructure-only migration

Phased Migration Roadmap:

- Multi-phase migration plans (typically 3-4 phases over 6 months)
- Service-by-service extraction prioritizing compliance-critical components
- Effort estimation (person-months), timeline, and deliverables per phase

Zensar Technologies

OCT 2025 AES – Diamond Project Team

Mentor : kaushik.saha@zensar.com

: sameerkumar.sharma@zensar.com

- Dependency tracking and risk assessment for each phase
- Breaking cyclic dependencies with event-driven architecture

Demographic Data Migration:

- Security-first approach for sensitive data (SSN, credit cards, PII)
- Encryption strategies (AES-256-GCM with cloud KMS, PCI-DSS tokenization)
- Data isolation patterns (dedicated PII vault microservices)
- Compliance features for GDPR, HIPAA, PCI-DSS with code transformation examples

Cost-Benefit Analysis:

- POD annual operating cost estimation (on-premise infrastructure, licenses, compliance risk)
- POA annual operating cost projection (cloud services, pay-as-you-go pricing)
- Migration investment calculation (development effort, tools, training)
- ROI break-even analysis and annual savings projections

Risk Management:

- Automated risk identification across security, compliance, data loss, performance, downtime
- Impact assessment (CRITICAL, HIGH, MEDIUM, LOW) with mitigation strategies
- Zero-downtime migration strategies using gradual traffic shifting

AI Model Support:

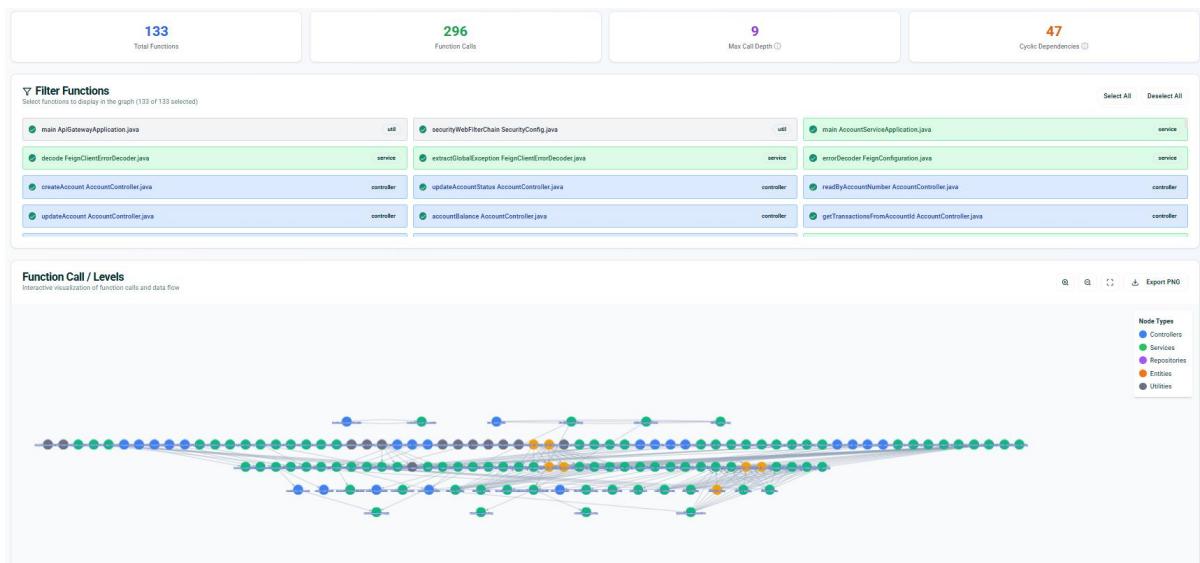
- OpenAI GPT-4o for strategic migration planning and architecture recommendations
- Ollama local models (Code Llama, Deepseek Coder, Mistral) for privacy-first offline analysis
- HuggingFace CodeBERT for code pattern matching and transformation suggestions
- Multi-AI approach combining strategic planning with tactical execution

13. Enterprise Security & Authentication

- Database-based user authentication with bcrypt password encryption
- Secure session management with express-session
- Multi-tenant support with user-based project association
- PostgreSQL database with Drizzle ORM
- Role-based access control

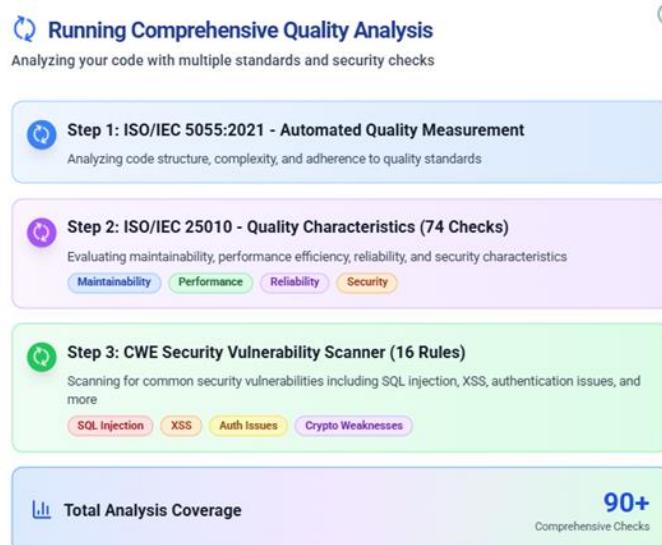
14. GitHub Integration

- Direct repository import via GitHub URL
- Automatic repository cloning and analysis
- Version-controlled project tracking
- Support for public and private repositories



15. Data Image and Quality Measures

Comprehensive code quality and data lineage analysis aligned with ISO/IEC 25010 standards:



ISO-5055 Quality Metrics:

- **Reliability:** Defect density, error-prone constructs, exception handling coverage
- **Security:** CWE (Common Weakness Enumeration) vulnerability detection
- **Maintainability:** Cyclomatic complexity, code duplication, modularity scores
- **Performance Efficiency:** Algorithm complexity analysis, resource utilization patterns
- **Testability:** Test coverage gaps, testable code patterns

Data Lineage & Impact Analysis:

- **Data Flow Mapping:** Track how demographic/sensitive data flows through the codebase
- **Field-to-Function Mapping:** Identify which functions process specific PII/PHI fields

- **Cross-Module Dependencies:** Understand data propagation across service boundaries
- **Database-to-UI Traceability:** Complete data journey from database to user interface
- **Compliance Impact:** Assess GDPR/HIPAA/PCI-DSS compliance at the data flow level

Quality Visualization:

- **Heat Maps:** Visual representation of code quality metrics across modules
- **Dependency Graphs:** Complex relationship mapping with quality scoring
- **Technical Debt Dashboard:** Prioritized list of improvement opportunities
- **Quality Trends:** Track code quality improvements over time (multi-version analysis)

Actionable Insights:

- Automated quality gate recommendations (e.g., "Do not deploy with Reliability Score < 80%")
- Risk-based prioritization of refactoring efforts
- Cost estimation for technical debt remediation
- Best practice alignment suggestions

Enterprise Add-on - Under Development

16. Advanced Intelligence Features ([Enterprise Add-on - Under Development](#))

Next-generation capabilities delivering unprecedented value for enterprise code management:

Real-Time Collaboration Intelligence

- Multi-user collaborative code analysis
- Live annotations and synced review sessions
- Quick version comparison and team activity view

Predictive Analytics Engine

- Technical debt & defect risk forecasting
- Maintenance cost and quality trend prediction
- Optimized resource/refactoring recommendations

Automated Code Transformation

- One-click AI refactoring with safe rollback
- Auto-apply SOLID/design patterns & templates
- Batch modernization across modules

Continuous Code Intelligence (CI/CD Integration)

- Auto-analysis on commits/PRs with quality gates
- Regression alerts and live code health metrics
- Automated weekly/monthly reports

Multi-Repository Portfolio Analysis

- Cross-project quality & pattern comparison
- Duplicate/reusable code detection
- Enterprise-level compliance and portfolio health

Advanced Security & Compliance Intelligence

- SAST/DAST integration with AI fix suggestions
- CVE monitoring and vulnerability remediation
- GDPR/HIPAA/ISO compliance insights

Smart Dependency Management

- Vulnerability and license compliance scanning

- Safe upgrade path & risk scoring
- Secure alternative dependency suggestions

Custom Pattern Builder (No-Code)

- Drag-and-drop pattern creation
- AI-assisted regex/pattern design
- Sandbox validation and reusable library

Executive Intelligence Dashboard

- Technical debt, quality & compliance KPIs
- Productivity, ROI & risk heat maps
- AI-driven strategic recommendations

AI-Powered Code Generation

- Auto-generate DTOs, APIs, tests, docs
- Intelligent code completion & optimization
- Automated API specs & migration helpers

Enterprise Integration Ecosystem

- Jira/ServiceNow ticket automation
- Slack/Teams/GitHub integration
- Auto-publish docs to Confluence/SharePoint

Vector Database Intelligence

- Semantic code search & clone detection
- Change impact & technical debt clustering
- Compliance/PII pattern mapping

Generative AI Functions

- Code explanation, migration & test generation
- Optimization & security patch suggestions
- ETL/data migration & compliance templates

Agentic Orchestrator with MCP

- Multi-agent coordination with shared context
- Automated workflow chaining & smart routing
- Full traceability, error recovery & observability

Value Proposition:

- **90% Faster Decision Making:** Real-time insights vs. weeks of manual analysis
- **50% Reduction in Production Defects:** Predictive analytics catch issues early
- **60% Faster Onboarding:** New developers understand codebase in days vs. weeks
- **\$1M+ Annual Savings:** Combined savings from efficiency, risk reduction, faster time-to-market

Architecture Overview

Frontend Architecture

- Component-based React application with TypeScript
- Tailwind CSS with shadcn/ui for consistent design
- TanStack Query for server state management
- Wouter for client-side routing
- Three-phase UX: Upload → Processing → Results

Backend Architecture

- Modular Express.js server with TypeScript

Zensar Technologies

OCT 2025 AES – Diamond Project Team

Mentor : kaushik.saha@zensar.com

: sameerkumar.sharma@zensar.com

- RESTful API design with Swagger documentation
- Service layer pattern (demographicScanner, demographicClassAnalyzer, swaggerGenerator)
- Storage abstraction layer (in-memory or PostgreSQL)
- File processing pipeline (ZIP → Extract → Parse → Analyze)

Data Storage

- Flexible storage abstraction (IStorage interface)
- Current: In-memory storage (MemStorage)
- Future: PostgreSQL with Drizzle ORM
- JSON format for analysis results
- User authentication and project metadata persistence

AI Analysis Pipeline

1. **Code Extraction:** ZIP/GitHub → Source files
2. **Pattern Matching:** Regex/Excel field detection
3. **Class Analysis:** Function extraction and demographic field mapping
4. **AI Processing:** Local LLM or OpenAI GPT-4o analysis
5. **Report Generation:** PDF/DOC/HTML with diagrams

Future Roadmap

- Enhanced ML models for field suggestion accuracy
- Real-time collaboration features
- Integration with CI/CD pipelines
- Advanced code refactoring suggestions
- Performance optimization for 10,000+ file projects
- Custom demographic pattern builder UI
- Multi-repository comparative analysis