

euGENE

Biomedical Knowledge Graph & Agentic AI Platform

ENTERPRISE PROJECT VALIDATION DOCUMENT

Version 1.0 | March 2026 Prepared for CSL Behring | Confidential

Table of Contents

1. Project Overview & Business Purpose
2. Detailed Architecture Diagram
3. Technology Stack & Dependencies
4. Database Connectivity & Architecture
5. Component / Service Inventory
6. API Endpoint Catalog
7. MCP Tooling & Agent Skills Design
8. Authentication & Security Architecture
9. Prompt Engineering & LLM Configuration
10. Workflow Orchestration
11. Deployment Model & CI/CD
12. Environment Configuration Matrix
13. Monitoring, Observability & Health Checks
14. Testing Strategy & Quality Assurance
15. Data Ingestion & Knowledge Strategy
16. Project Structure & Folder Hierarchy
17. Development Guide & Setup Instructions
18. Risks, Mitigations & Open Questions
- Appendix A: Complete Cypher Query Catalog
- Appendix B: Environment Variable Reference
- Appendix C: Release Notes

1. Project Overview & Business Purpose

1.1 What the System Does

euGENE is a biomedical knowledge graph and agentic AI platform developed for CSL Behring. It enables competitive intelligence across the pharmaceutical landscape by integrating data from DrugBank, ClinicalTrials.gov, USPTO, PubMed, and internal CSL sources into a Neo4j graph database. Users can explore drug pipelines, disease-gene relationships, patent landscapes, and organizational partnerships through both a structured REST API and a conversational AI chat interface powered by LLM agents (Anthropic Claude and OpenAI GPT).

1.2 Target Audience & User Personas

1.3 Key Business Capabilities

- Drug-Disease Intelligence: Explore indications, contraindications, off-label uses, drug aliases
- Gene-Protein Analysis: Map drug-protein targets, disease-protein associations, pathway relationships
- Patent Landscape: Search patents by drug, clinical trial, or gene/protein with pagination
- Organizational Assets: Discover company drug pipelines, clinical trial sponsorships, partnerships
- PubMed Integration: Link published research to drugs, diseases, and genes
- Agentic AI Querying: Natural language questions answered via ReAct agent with 12 graph tools
- Graph Analytics: Node similarity (cosine), community detection (Leiden), GraphSAGE embeddings

1.4 Quantitative Summary

2. Detailed Architecture Diagram

2.1 System Architecture (Five-Tier)

```
+=====+
|| PRESENTATION LAYER (Tier 5) ||
+=====+

|| || | | | |
|| +-----+ +-----+ ||
|| | eugene-agent-ui (Streamlit) | | Neo4j Browser | ||
|| | Port: 18501 (local) / 8501 | | Port: 17474 (local) / 7474 | ||
|| | | | | ||
|| | - Chat message interface | | - Graph visualization | ||
|| | - Suggestion pills | | - Cypher query console | ||
|| | - Tool selection sidebar | | - Schema browser | ||
|| | - OAuth token input | | | ||
|| | - Streaming response display | | | ||
|| +-----+ +-----+ ||
|| | ||
+=====+

| HTTP POST /agent/api/query/stream
| Authorization: Bearer <EUGENE_JWT>
v
+=====+
|| AGENT ORCHESTRATION LAYER (Tier 4) ||
+=====+

|| || | | | | | |
|| +-----+ ||
|| | eugene-agent-ws (FastAPI + Strands Framework) |||
|| | Port: 18001 (local) / 8000 |||
|| | |||
|| | +-----+ +-----+ |||
|| | | LLM Factory | | Agent Engine | |||
|| | | - Anthropic Claude Sonnet | | - ReAct Pattern | |||
|| | | - OpenAI GPT-4.1-mini | | - 10-msg sliding window | |||
|| | | - Auto-detection | | - FileSessionManager | |||
|| | +-----+ +-----+ |||
|| | |||
|| | Local Tools: calculator, current_time, python_repl, http_request |||
|| +-----+ ||
|| | ||
+=====+

| MCP HTTP Streaming + JWT
v
+=====+
|| MCP TOOL SERVER LAYER (Tier 3) ||
+=====+

|| ||
```

```

|| +-----+|| | | | | | | | |
|| | eugene-mcp (FastMCP + JWT Verifier) |||
|| | Port: 18443 (local) / 8000 |||
|| | |||
|| | +-----+ +-----+ +-----+ |||
|| | | Identity Tools (1) | | Fetch Tools (2) | | Node Tools (2) | |||
|| | | - fetch_identity | | - fetch_by_label | | - lookup_node | |||
|| | | | - fetch_similar | | - fetch_details | |||
|| | +-----+ +-----+ +-----+ |||
|| | +-----+ +-----+ +-----+ |||
|| | | Drug Tools (1) | | Fact Tools (1) | | Graph Tools (3) | |||
|| | | - fetch_aliases | | - fetch_facts | | - relationships | |||
|| | | | | - paths, reachable | |||
|| | +-----+ +-----+ +-----+ |||
|| | +-----+ |||
|| | | Org Tools (2) | |||
|| | | - find_org_names | |||
|| | | - find_org_assets | |||
|| | +-----+ |||
|| +-----+||
|| | ||
+=====+
| HTTP REST + JWT
v
+=====+
|| CORE API LAYER (Tier 2) ||
+=====+
|| || | | | | | | | |
|| +-----+||
|| | eugene_ws (FastAPI + Uvicorn) |||
|| | Port: 18000 (local) / 8000 |||
|| | |||
|| | Routers (20): |||
|| | auth | health | stats | labels | count | node_lookup | node_details |||
|| | n_hop | search_path | facet | similarity | drug_aliases |||
|| | patent_search | patent_count | pubmed_search | pubmed_count |||
|| | facts | organization_search | root | release_notes |||
|| | |||
|| | Architecture: DDD + Hexagonal (model/provider/mapper/adaptor/conf) |||
|| +-----+||
|| | ||
+=====+
| Bolt Protocol (neo4j:// or bolt+ssc://)
v
+=====+
|| DATA PERSISTENCE LAYER (Tier 1) ||
+=====+
|| || | | | |
|| +-----+ +-----+ ||
|| | Neo4j 5.26.9 Community Edition | | Milvus (Vector DB) | ||
|| | Port: 17687 (Bolt) / 17474 (HTTP) | | milvus-lite (embedded) | ||
|| | | | ||
|| | Plugins: | | Model: minishlab/potion-base-8M| ||
|| | APOC 5.25.0 Extended | | Purpose: Semantic search | ||
|| | Graph Data Science (GDS) 2.13.2+ | | embeddings for drugs, | ||
|| | | diseases, documents | ||
|| | Data: | | | ||
|| | 129,375 nodes | | | ||
|| | 4,050,249 relationships | | | ||

```

```
|| | 4,567,749 properties | | | ||
|| +-----+ +-----+ ||
|| ||
+=====+
```

2.2 Deployment Environments

```
DEPLOYMENT TARGETS:
+-----+ +-----+ +-----+
| LOCAL DEVELOPMENT | | DIFFLABS (Dev) | | AIA (Staging/Prod) |
| | | | |
| Docker Compose | | AWS us-east-1 | | AWS eu-central-1 |
| 5 containers | | ECS Fargate | | ECS Fargate |
| Ports: 17xxx/18xxx | | ALB + Target Groups | | ALB + Target Groups |
| Neo4j: local volume | | EC2 r7i.xlarge | | EC2 r7i.xlarge |
| Auth: bypass mode | | Acct: 087084717211 | | Acct: 010928221940 |
+-----+ +-----+ +-----+

DATA FLOW:
User ---> Streamlit UI ---> Agent WS ---> MCP Server ---> Core API ---> Neo4j
| (8501) (8001) (8443) (8000) (7687)
| |
+--- JWT Token forwarded through entire chain (HS256) -----+
```

2.3 Authentication Flow

```
PRODUCTION AUTH FLOW:
+-----+ +-----+ +-----+ +-----+
| User | --> | /login | --> | Microsoft | --> | Entra ID |
| | | endpoint | | Entra ID | | ID Token |
+-----+ +-----+ | (OAuth 2.0) | | (RS256) |
+-----+ +-----+
|
JWKS validation |
(8h cached keys) v
+-----+
| Eugene |
| JWT |
| (HS256) |
| 24h TTL |
+-----+
|
Token forwarded: UI -> Agent WS -> MCP -> API |
Each service validates independently v

LOCAL DEV AUTH FLOW:
+-----+ +-----+ +-----+
| User | --> | /login | --> | Dev Token |
| | | (local) | | Immediate |
+-----+ +-----+ | HS256 |
+-----+
User: eugene.test@cs1hering.com
Roles: [user.public.read]
```

3. Technology Stack & Dependencies

3.1 Core Technology Matrix

3.2 Key Python Dependencies (requirements.txt)

4. Database Connectivity & Architecture

4.1 Connection Configuration

4.2 Neo4j Server Configuration (Production)

4.3 Node Label Inventory

4.4 Relationship Type Inventory

5. Component / Service Inventory

5.1 Service: neo4j (Data Tier)

5.2 Service: eugene_ws (Core API Tier)

5.3 Service: eugene_mcp (MCP Tool Tier)

5.4 Service: eugene_agent_ws (Agent Tier)

5.5 Service: eugene_agent_ui (Frontend Tier)

6. API Endpoint Catalog

6.1 Core API Endpoints (eugene_ws) - 38 Total

6.2 Agent API Endpoints (eugene_agent_ws)

7. MCP Tooling & Agent Skills Design

7.1 MCP Tool Inventory (12 Tools)

7.2 Future Tools (Planned, Currently Disabled)

8. Authentication & Security Architecture

8.1 Three-Layer Authentication

8.2 JWT Token Structure

```
Eugene JWT Claims:
{
  "tid": "f8645748-68c6-4eec-bd61-c71341a6ed7d", // Tenant ID
  "sub": "<user-oid>", // Subject (user OID)
  "upn": "user@cslbehrling.com", // User Principal Name
  "roles": ["user.public.read"], // Assigned roles
  "iat": 1711500000, // Issued at
  "exp": 1711586400, // Expires (24h later)
  "iss": "https://eugene.ai.cslg1.cslg.net/<tid>", // Issuer
```

```
"aud": "api://eugene/<client_id>" // Audience
}
```

8.3 Role-Based Access Control

8.4 Security Posture Assessment

9. Prompt Engineering & LLM Configuration

9.1 Agent System Prompt

```
SYSTEM_PROMPT = """
You are an agent tasked with helping investigate biomedical
companies to assess competitive threat and collaboration
opportunities. In our case assets here mean any drug, disease,
patents, clinical trials, intellectual property, or financial
deals the company may have involvement. As an agent follow the
Reason, Act, Observe (ReAct) pattern. You are an agent that
may call tools to retrieve data.
"""
```

9.2 LLM Provider Configuration

Provider auto-detection priority: 1) LLM_PROVIDER env var, 2) ANTHROPIC_API_KEY present -> Anthropic, 3) OPENAI_API_KEY present -> OpenAI. OpenAI client uses max_retries=3, timeout=60s.

9.3 Conversation Management

- Sliding window: 10 messages retained per conversation
- Per-turn truncation: Max 2 result items per tool response
- Session persistence: FileSessionManager (local disk)
- Conversation ID: UUID-based tracking across turns
- Prompt validation: Max 2048 chars, blocks special chars (% _ \$; : ^ * @)

9.4 Suggestion Pills (UI Pre-built Queries)

10. Workflow Orchestration

10.1 Query Execution Flow (Streaming)

1. User submits prompt via Streamlit UI
2. UI sends POST /agent/api/query/stream
Headers: Authorization: Bearer <EUGENE_JWT>
Body: {prompt, conversation_id, include_tools}
3. Agent WS validates JWT + checks user in allowlist
4. Agent WS initializes EugeneDataAgent:
 - a. LlmFactory creates model (Anthropic or OpenAI)
 - b. MCPClient connects to MCP server with JWT
 - c. MCPClient.list_tools() discovers 12 available tools
 - d. Local tools added (calculator, current_time, python_repl)
 - e. Strands Agent created with system prompt + all tools
5. Agent executes ReAct loop:
 - a. REASON: LLM analyzes query, plans approach
 - b. ACT: LLM calls selected MCP tool(s)
 - c. MCP tool authenticates with Core API (JWT forwarded)
 - d. Core API executes parameterized Cypher query
 - e. OBSERVE: LLM reviews results, decides next action
 - f. Repeat until answer is complete
6. Streaming events emitted:

```
{"type": "session", "session_id": "<uuid>"}
{"type": "content", "content": "token..."}
{"type": "done", "content": ""}
7. UI displays tokens in real-time with cursor indicator
8. Metrics logged: total_tokens, execution_time, tools_used
```

10.2 Docker Compose Startup Orchestration

Dependency Chain (automatic via depends_on + healthcheck):

```
neo4j (healthcheck: wget localhost:7474, 10s interval, 30s startup)
|-> eugene_ws (depends: neo4j HEALTHY)
| (healthcheck: urllib localhost:8000/health, 10s, 15s startup)
|-> eugene_mcp (depends: eugene_ws HEALTHY)
|-> eugene_agent_ws (depends: eugene_mcp)
|-> eugene_agent_ui (depends: eugene_agent_ws)
```

11. Deployment Model & CI/CD

11.1 CI/CD Pipeline (GitLab)

11.2 Branch Strategy

```
feature/* -> Docker build only (no deploy)
qa -> Full pipeline: build + scan + terraform plan/apply (staging)
main -> Full pipeline: build + scan + terraform plan/apply (production)

AWS OIDC Authentication:
GitLab OIDC Token -> AWS STS assume-role-with-web-identity
Role: arn:aws:iam::010928221940:role/csl-gitlab-ci-cd-scop-all-branches
Session duration: 3600 seconds
```

11.3 Docker Images

11.4 Terraform Modules

12. Environment Configuration Matrix

13. Monitoring, Observability & Health Checks

13.1 Health Check Endpoints

13.2 API Canaries (Automated Monitoring)

Canary targets: difflabs ALB (internal-eugene-search-alb-616664632.us-east-1.elb.amazonaws.com), AIA Staging (eugene-staging.ai.cslg1.cslg.net), AIA Prod (eugene.ai.cslg1.cslg.net).

13.3 Agent Execution Metrics

- total_tokens: Total LLM tokens consumed per query
- execution_time: Total agent execution time (seconds, 4 decimal precision)
- tools_used: List of unique tool names invoked during query
- cycle_durations: Time per ReAct loop iteration
- @log_time decorator applied to agent init, execute, and stream functions

14. Testing Strategy & Quality Assurance

14.1 Test Framework Configuration

14.2 Pre-Commit Hooks

- Terraform validation (difflabs/iac/ modules)
- Black code formatter on src/
- pytest full suite execution
- Exit code 0 (pass) or 1 (fail, blocks commit)

14.3 Recommended Testing Expansion

15. Data Ingestion & Knowledge Strategy

15.1 Data Sources & Status

15.2 Three-Phase Ingestion Pipeline

```
PHASE 1: BULK IMPORT (Foundational)
sudo systemctl stop neo4j.service
neo4j-admin database import full \
--nodes=nodes.csv \
--relationships=edges_dedup.csv \
--overwrite-destination --verbose
# Result: 129,375 nodes, 4,050,249 relationships in 7.6 seconds
# Peak memory: 1.034 GB
# Source: s3://knowledge-graph-external-data/
sudo systemctl start neo4j.service

PHASE 2: FEATURE ENRICHMENT
python src/batch_load_disease_features.py # 44,133 features, ~53 min
python src/batch_load_drug_features.py # 7,957 features, ~9.5 min
# Uses APOC batch transactions (1000 rows/batch)

PHASE 3: SUPPLEMENTARY DATA
# Clinical trials: eugene/saved_queries/ingest_csv/clinical_trials.cypher
# Patents: bin/docker/ingest-bio-data.py (USPTO API download)
# PubMed: src/pubmed/ module (download + APOC loading)
```

15.3 Local Sample Data Ingestion

```
python bin/docker/ingest-sample-data.py
# Connects: bolt://localhost:17687
# Loads: 6 data sources (pharma data, gene editing, entities, relationships, orgs)
# Uses: MD5-based ID generation, MERGE operations
```

16. Project Structure & Folder Hierarchy

```
knowledgeGraph/
+-- src/ # Core API source code
| +-- eugene_ws.py # FastAPI application entry point
| +-- router/ # API routing layer
| | +-- auth/ # Authentication (Entra ID, JWT, RBAC)
| | +-- health_router.py # Health check endpoint
| | +-- root_router.py # Root welcome endpoint
| | +-- release_notes_router.py
| +-- foundation/ # Core graph domain
| | +-- model/ # Node/Relationship enums, response models
| | +-- provider/ # Business logic (n-hop, path, facts)
| | +-- mapper/ # Data transformers (graph, facts, patents)
| | +-- router/ # API endpoints (labels, count, node, graph, etc.)
| | +-- infra/db/adapters/ # Neo4j Cypher query adapters
| | +-- conf/ # Dependency injection factory
```



```

| +-- organization/ # Organization domain
| +-- patent/ # Patent domain (pgpub, application)
| +-- pubmed/ # PubMed domain
| +-- clinicaltrail/ # Clinical trial domain
| +-- tpp/ # Target Product Profile domain
| +-- graph/ # Graph analytics & community detection
| +-- document/ # Document analysis & parsing
| +-- centree/ # Drug development projects
| +-- stats/ # Database statistics
| +-- infra/ # Infrastructure (LLM, embeddings, DB connection)
| +-- util/ # Utilities (regex, timer, case helper)
+-- agents/ # Agent microservices
| +-- eugene-mcp/ # MCP Tool Server (FastMCP)
| | +-- src/tools/ # 7 tool classes (12 tools)
| | +-- src/auth/ # JWT verifier
| | +-- src/conf/ # Configuration
| +-- eugene-agent-ws/ # Agent Backend (Strands)
| | +-- src/query/agent/ # EugeneDataAgent (ReAct)
| | +-- src/query/infra/llm/ # LLM Factory (Claude, GPT)
| | +-- src/query/router/ # Agent API endpoints
| | +-- src/query/model/ # Request/Response models
| +-- eugene-agent-ui/ # Chat UI (Streamlit)
| +-- src/ # UI application
+-- docker/ # Local Dockerfiles
+-- containers/ # Production Dockerfiles (AIA)
+-- difflabs/ # Dev environment (containers + Terraform)
+-- infrastructure/ # Production Terraform (AIA)
| +-- modules/eugene-ec2/ # EC2 module
| +-- modules/eugene-services/ # ECS services module
| +-- environments/ # qa-ec2, qa-services configs
+-- eugene/ # Neo4j setup, notebooks, saved queries
+-- docs/ # Documentation
+-- bin/ # Operational scripts
| +-- docker/ # local-up, local-down, ingest scripts
| +-- docker/ecr/ # ECR login, build, tag, push
+-- api-canaries/ # API health monitoring
+-- tests/ # Test data directory
+-- docker-compose.yml # Local 5-service orchestration
+-- docker.env.template # Environment variable template
+-- requirements.txt # Python dependencies
+-- .gitlab-ci.yml # CI/CD pipeline
+-- pyproject.toml # Project metadata
+-- .pytest.ini # Test configuration

```

17. Development Guide & Setup Instructions

17.1 Prerequisites

- Python 3.12+ (3.13 recommended)
- Docker Desktop (latest)
- Git
- Terraform >= 1.0.0 (for infrastructure work)
- AWS CLI v2 (for ECR push / S3 access)

17.2 Local Development Setup

```

# 1. Clone repository
git clone <repo-url>
cd knowledgeGraph

# 2. Configure environment

```

```

cp docker.env.template docker.env
# Edit docker.env: set API keys, Neo4j password

# 3. Start all services
docker compose up --build
# Or use helper script:
bin/docker/local-up.sh --build

# 4. Ingest sample data (optional)
python bin/docker/ingest-sample-data.py

# 5. Access services:
# Neo4j Browser: http://localhost:17474
# Core API Swagger: http://localhost:18000/docs
# Core API Login: http://localhost:18000/login
# Agent API Docs: http://localhost:18001/docs
# Chat UI: http://localhost:18501

# 6. Stop services
bin/docker/local-down.sh # Keep data
bin/docker/local-down.sh --reset # Delete Neo4j volumes

```

17.3 Running Core API Locally (without Docker)

```

python -m venv venv_eugene_ws
source venv_eugene_ws/bin/activate
pip install -r requirements.txt
cp .env.template .env
# Edit .env with local Neo4j credentials
uvicorn eugene_ws:app --app-dir src --host 127.0.0.1 --port 8000 --reload

```

17.4 Running Tests

```

pytest # All tests with coverage
pytest -n 4 # 4 parallel workers
pytest -m integration # Integration tests only
pytest -m serial # Serial (DB-dependent) tests
pytest src/organization/ # Single domain tests

```

17.5 Deploying to difflabs

```

# Login to ECR
bin/docker/ecr/login.sh

# Build, tag, push images
bin/docker/ecr/build.sh
bin/docker/ecr/tag.sh
bin/docker/ecr/push.sh

# Deploy infrastructure
cd difflabs/iac/eugene-services
./bin/init-difflabs.sh
./bin/plan-difflabs.sh
./bin/apply-difflabs.sh

```

18. Risks, Mitigations & Open Questions

18.1 Risk Register

18.2 Open Questions

Appendix A: Complete Cypher Query Catalog

Count All Nodes by Label

```
MATCH (n:`{label}`)
RETURN count(n) as count
```

Find Node by Value

```
MATCH (n)
WHERE n.node_name =~ '(?i).*{value}.*'
AND (n.is_hidden IS NULL OR NOT n.is_hidden)
RETURN labels(n), n.node_id, n.node_name
```

Node Details by IDs

```
MATCH (n)
WHERE (n.is_hidden IS NULL OR NOT n.is_hidden)
AND n.node_id IN $node_ids
RETURN labels(n) as labels, n.node_id as id,
n.node_name as value, n.node_source as source,
n.description as description
```

N-Hop Graph Traversal

```
MATCH (start {node_id: $start_id})-[r]-{0,{n_hop}}(end)
UNWIND(r) AS rel
RETURN DISTINCT startNode(rel).node_name,
endNode(rel).node_name, type(rel)
ORDER BY startName SKIP $offset LIMIT $limit
```

Drug Aliases by Name

```
MATCH (n) WHERE n.node_name = $drug_name
AND (n.is_hidden IS NULL OR NOT n.is_hidden)
CALL apoc.path.subgraphNodes([n],
{relationshipFilter: 'has_drug_alias'}) YIELD node
RETURN DISTINCT node.node_name, node.node_id,
node.drug_bank_id, 'drug' in labels(node) as is_canonical
```

Patents by Drug

```
MATCH (n:`DRUG`)-[r:`disclosed_in`]-(n2:`Patent_Application`)
WHERE n.node_id = $drug_id
RETURN n.node_id, n.node_name, n2.patent_id
ORDER BY patent_id DESC SKIP $offset LIMIT $limit
```

PubMed by Gene

```
MATCH (n:`GENE_PROTEIN`)-[r:`analyzed_in`]-(n2:`Research`)
WHERE n.node_name = $gene
RETURN n2.pmid, n2.title
ORDER BY n2.pmid DESC SKIP $offset LIMIT $limit
```

Organization Search

```
MATCH (o:`Organization`)
WHERE o.organization_canonical_name =~ $pattern
RETURN o.org_id, o.organization_canonical_name,
o.organization_type
SKIP $skip LIMIT $limit
```

Organization Assets

```
MATCH (start:`Organization`)-[rel1]-(ct:`ClinicalTrial`)
-[rel2]-(end:`DISEASE`|`DRUG`)
WHERE start.org_id = $org_id
RETURN start.organization_canonical_name,
ct.nct_id, labels(end), end.node_name
ORDER BY trial DESC SKIP $skip LIMIT $limit
```

GDS Node Similarity

```
MATCH (d1:`DRUG`|`DISEASE`)
WHERE d1.node_name =~ '(?i).*{value}.*'
```

```
CALL gds.nodeSimilarity.filtered.stream(
  '{projection}', {
    degreeCutoff: 1, similarityCutoff: .45,
    similarityMetric: 'COSINE',
    sourceNodeFilter: [d1]
  }) YIELD node1, node2, similarity
RETURN similarity, gds.util.asNode(node1).node_name,
gds.util.asNode(node2).node_name
ORDER BY similarity DESC
```

Appendix B: Environment Variable Reference

Appendix C: Release Notes

--- End of Document ---
Generated from complete codebase analysis | euGENE Knowledge Graph | March 2026

Persona	Use Case	Interface
R&D Scientists	Explore drug-disease-gene relationships, find competitive drugs	Chat UI, API
Patent Analysts	Search patent landscapes, find patent-drug linkages	API, Chat UI
Business Development	Assess organizational assets, competitive threat analysis	Chat UI
Data Engineers	Ingest new datasets, manage graph schema, run analytics	CLI, API
DevOps / SRE	Deploy, monitor, scale the platform	Terraform, CI/CD, CloudWatch

Metric	Value
Graph Nodes	129,375
Graph Relationships	4,050,249
Graph Properties	4,567,749
Node Types (Labels)	36 distinct types
Relationship Types	27 distinct types
Core API Endpoints	38
MCP Agent Tools	12
Microservices	5 (Neo4j, Core API, MCP, Agent WS, Agent UI)
LLM Providers Supported	2 (Anthropic Claude, OpenAI GPT)
AWS Environments	2 (difflabs us-east-1, AIA eu-central-1)
Authentication Layers	3 (Entra ID, Eugene JWT, Agent Allowlist)

Layer	Technology	Version	Purpose
Runtime	Python	3.12-3.13	Application runtime
API Framework	FastAPI	Latest (via pip)	REST API with auto OpenAPI docs
ASGI Server	Uvicorn	Latest	Production HTTP server
Graph Database	Neo4j Community	5.26.9	Knowledge graph storage
Neo4j Driver	neo4j (Python)	5.25.0	Database connectivity
Graph Plugins	APOC Extended	5.25.0	Utility procedures (path, meta, merge)
Graph Plugins	Graph Data Science	2.13.2+	ML algorithms (similarity, embeddings)
Vector DB	Milvus Lite	2.4.10	Semantic search embeddings
MCP Framework	FastMCP	3.0.0b1	Model Context Protocol server
Agent Framework	Strands Agents	1.21.0	LLM agent orchestration
UI Framework	Streamlit	1.52.2	Chat web interface
LLM (Anthropic)	Claude Sonnet 4	claude-sonnet-4-20250514	Primary LLM provider
LLM (OpenAI)	GPT-4.1-mini	gpt-4.1-mini	Secondary LLM provider
Embeddings	potion-base-8M	minishlab/potion-base-8M	Semantic vector embeddings
Auth (Enterprise)	MSAL	Latest	Microsoft Entra ID OAuth 2.0
Auth (JWT)	PyJWT / python-jose	Latest	HS256 token signing/validation

Layer	Technology	Version	Purpose
IaC	Terraform	>= 1.0.0	AWS infrastructure provisioning
CI/CD	GitLab CI	Cloud hosted	Build, scan, deploy pipelines
Containers	Docker / Docker Compose	Latest	Service containerization
Cloud	AWS (ECS, EC2, ECR, S3, ALB)	Latest	Production hosting
Data Processing	Pandas	2.2.3	DataFrame operations
ML/NLP	Transformers	4.46.2	NLP model loading
ML/NLP	Sentence-Transformers	3.3.0	Embedding models
ML/NLP	PyTorch (CPU)	Latest	ML framework (CPU-only)

Category	Packages
Web Framework	fastapi, uvicorn, starlette, pydantic==2.8.2, pydantic-settings==2.5.2
Database	neo4j==5.25.0, pymilvus==2.4.3, milvus-lite==2.4.10, kuzu==0.6.0
LLM / LangChain	langchain==0.3.1, langchain-aws==0.2.1, langchain-openai==0.2.1, openai==1.42.0
ML / NLP	transformers==4.46.2, sentence-transformers==3.3.0, model2vec==0.3.2, torch (CPU)
Data Science	pandas==2.2.3, numpy==1.26.4, scikit-learn, scipy, networkx, graspologic
Auth	msal, PyJWT, python-jose, cryptography
AWS	boto3==1.34.162, botocore==1.34.162
HTTP	httpx==0.27.0, requests==2.32.3
Utilities	python-dotenv==1.0.1, orjson==3.10.7, regex==2024.7.24, click, Rich
Dev Tools	black, pytest, pytest-cov, pytest-xdist, ruff, mypy

Environment	URI	TLS	Auth
Local (Docker)	bolt://neo4j:7687 (internal) bolt://localhost:17687 (host)	Disabled	neo4j / eugene_local_2024
difflabs (Dev)	bolt+ssc://10.88.203.244:7687	Self-signed (EC P-384)	neo4j / <secret>
AIA (Prod)	neo4j+ssc://<ec2-ip>:7687	Self-signed (EC P-384)	neo4j / <AWS Secrets>

Setting	Value	Purpose
server.memory.heap.initial_size	16384m	JVM heap (16 GB)
server.memory.heap.max_size	16384m	JVM max heap (16 GB)
server.memory.pagecache.size	512M	Page cache for disk reads
server.bolt.tls_level	REQUIRED	Enforce TLS on Bolt connections
server.default_listen_address	0.0.0.0	Accept connections from any host
dbms.security.procedures.unrestricted	apoc.*, gds.*	Allow all APOC and GDS procedures
dbms.cypher.lenient_create_relationship	true	Allow flexible relationship creation
db.tx_log.rotation.retention_policy	2 days 2G	Transaction log retention
JVM Additional	--add-modules=jdk.incubator.vector	Enable Java vector API

ID	Label	Count	Source	Key Properties
8	drug	4,600	DrugBank	node_id, node_name, drug_bank_id, description
7	disease	17,100	Disease Ontology	node_id, node_name, MONDO_ID, MONDO_NAME
12	gene_protein	27,700	Protein DBs	node_id, node_name, annotation
4	ClinicalTrial	49,300	ClinicalTrials.gov	node_id, node_name, nct_id
32	Organization	26,100	Company/Sponsor	org_id, organization_canonical_name, organization_type
1	anatomy	Various	Anatomical Ontology	node_id, node_name
2	biological_process	Various	Gene Ontology	node_id, node_name
15	pathway	Various	Pathway DBs	node_id, node_name
9	effect_phenotype	Various	Phenotype data	node_id, node_name
10	exposure	Various	Environmental data	node_id, node_name
14	molecular_function	Various	Gene Ontology	node_id, node_name
16	patent	Pending	USPTO	node_id, patent_id
31	Patent_Application	Pending	USPTO	patent_id, node_name
30	Approved_Patent	Pending	USPTO	patent_id, node_name
400	pubmed_document	Pending	PubMed	pmid, title, doi
300	csl_tpp	Partial	CSL Internal	node_id, therapeutic_area

ID	Label	Count	Source	Key Properties
21	drug_product	Various	DrugBank	node_id, node_name
22	drug_synonym	Various	DrugBank	node_id, node_name
100	summary (GraphRAG)	Various	Generated	node_id, content
200	uspto_application	Pending	USPTO	node_id, app_number

ID	Relationship	Source -> Target	Description
15	indication	Drug -> Disease	Drug is indicated for disease
6	contraindication	Drug -> Disease	Drug is contraindicated
1	off-label use	Drug -> Disease	Drug used off-label
13	drug_protein	Drug -> Gene/Protein	Drug targets protein
10	disease_protein	Disease -> Gene/Protein	Disease associated with protein
12	drug_effect	Drug -> Effect/Phenotype	Drug has pharmacological effect
11	drug_drug	Drug -> Drug	Drug-drug interactions
7	disease_disease	Disease -> Disease	Disease comorbidities
19	protein_protein	Protein -> Protein	Protein interactions
18	pathway_protein	Pathway -> Protein	Protein in pathway
3	anatomy_protein_present	Anatomy -> Protein	Protein present in tissue
2	anatomy_protein_absent	Anatomy -> Protein	Protein absent in tissue
9	disease_phenotype_positive	Disease -> Phenotype	Disease manifests phenotype
8	disease_phenotype_negative	Disease -> Phenotype	Disease lacks phenotype
14	exposure_disease	Exposure -> Disease	Environmental exposure link
20	has_drug_alias	Drug -> Drug Synonym	Drug name synonym
21	disclosed_in	Drug -> Patent	Drug disclosed in patent
22	supports_patent_application	Trial -> Patent	Trial supports patent
23	patent_app_target	Gene -> Patent	Gene targeted in patent
24	featured_in	Drug -> Research	Drug featured in publication
25	analyzed_in	Gene -> Research	Gene analyzed in publication
26	evaluated_in	Trial -> Research	Trial evaluated in pub
200	has_publication	Entity -> PubMed	Entity linked to PMID

Property	Value
Image	neo4j:5.26.9-community-bullseye
Container Name	eugene-neo4j
Ports	17474:7474 (HTTP Browser), 17687:7687 (Bolt Driver)
Plugins	APOC 5.25.0 Extended, GDS 2.13.2+
Volumes	eugene_neo4j_data:/data, eugene_neo4j_logs:/logs
Health Check	wget --spider http://localhost:7474 (10s interval, 30s startup)
Memory	512M pagecache, 512m-1024m heap (local) / 16GB heap (prod)

Property	Value
Framework	FastAPI + Uvicorn
Container Name	eugene-ws
Port	18000:8000
Source	src/eugene_ws.py
Routers	20 (auth, health, stats, labels, count, node_lookup, node_details, n_hop, search_path, facet, similarity, drug_aliases, patent_search, patent_count, pubmed_search, pubmed_count, facts, organization_search, root, release_notes)
Architecture	DDD + Hexagonal (model/provider/mapper/adapters/conf per domain)
Dependencies	neo4j service (healthy)
Health Check	urllib http://localhost:8000/health (10s interval, 15s startup)
Special	CPU-only PyTorch, HuggingFace model cache

Property	Value
Framework	FastMCP 3.0.0b1 + JWT Verifier
Container Name	eugene-mcp
Port	18443:8000
Source	agents/eugene-mcp/src/eugene_mcp.py
Tools	12 tools in 7 classes (Identity, Fetch, Node, Drug, Fact, Graph, Organization)
Auth	HS256 JWT verification middleware on every tool call
Transport	Stateless HTTP streaming

Property	Value
Dependencies	eugene_ws service (healthy)
HTTP Client	httpx, 30s timeout, verify=False (internal network)

Property	Value
Framework	FastAPI + Strands Agents 1.21.0
Container Name	eugene-agent-ws
Port	18001:8000
Source	agents/eugene-agent-ws/src/eugene_chat_ws.py
root_path	/agent/api
Endpoints	POST /query (sync), POST /query/stream (SSE)
LLM Support	Anthropic Claude (claude-sonnet-4-20250514), OpenAI GPT (gpt-4.1-mini)
Agent Pattern	ReAct (Reason, Act, Observe)
Conversation	SlidingWindowConversationManager(window_size=10, per_turn=2)
Session	FileSessionManager (local), future: S3SessionManager
Auth	JWT validation + user allowlist (EUGENE_AGENT_ALLOWLIST)
Local Tools	calculator, current_time, python_repl, http_request (optional)

Property	Value
Framework	Streamlit 1.52.2
Container Name	eugene-agent-ui
Port	18501:8501
Source	agents/eugene-agent-ui/src/eugene_agent_ui.py
Layout	Wide, sidebar expanded
Features	Chat history, streaming responses, suggestion pills, tool selection, OAuth token input
Dependencies	eugene_agent_ws service
API Endpoint	EUGENE_AGENT_API_URL or http://localhost:8000/agent/api/query

Method	Path	Auth	Description
GET	/login	No	Initiate OAuth / issue local dev token
GET	/auth/callback	No	OAuth redirect callback from Entra ID
GET	/auth/whoami	Yes	Return authenticated user claims
GET	/	Yes	Welcome message with user info
GET	/health	No	Health check (returns {status: OK})
GET	/releases	No	Release notes history (v1 - v2.2)
GET	/stats	Yes	Database statistics (node/rel counts)
GET	/count/{label}	Yes	Count nodes by label type
GET	/labels/{label}	Yes	List nodes by label (paginated, max 50)
GET	/node/find/{node_value}	Yes	Find node ID by name (fuzzy optional)
POST	/node/details	Yes	Get details for list of node IDs (max 50)
GET	/graph/relationship/start/{start_id}	Yes	N-hop subgraph from node (n_hop: 1-2)
GET	/graph/path/start/{start_id}/end/{end_id}	Yes	Shortest path between nodes (n_hop: 1-4)
GET	/graph/reachability/start/{start_id}/end/{end_id}	Yes	Check reachability (n_hop: 1-4)
GET	/graph/facts/start/{start_id}	Yes	English-language facts (paginated)
POST	/facet/{label}	Yes	Faceted search by label and values
POST	/similarity/{label}	Yes	Cosine similarity via GDS (drug/disease)
GET	/drugs/aliases/{drug_name}	Yes	Drug aliases by name (fuzzy optional)
GET	/drugs/aliases/id/{drug_id}	Yes	Drug aliases by DrugBank ID
GET	/patents/drugs	Yes	Patents by drug_id (paginated, max 100)
GET	/patents/clinicaltrials	Yes	Patents by nct_id (paginated)
GET	/patents/geneproteins	Yes	Patents by gene_protein name (paginated)
GET	/count/patents/drugs	Yes	Count patents by drug_id

Method	Path	Auth	Description
GET	/count/patents/clinicaltrials	Yes	Count patents by nct_id
GET	/count/patents/geneproteins	Yes	Count patents by gene_protein
GET	/pmids/drugs	Yes	PubMed articles by drug_id (paginated)
GET	/pmids/clinicaltrials	Yes	PubMed articles by nct_id (paginated)
GET	/pmids/geneproteins	Yes	PubMed articles by gene_protein (paginated)
GET	/count/pmids/drugs	Yes	Count PubMed articles by drug_id
GET	/count/pmids/clinicaltrials	Yes	Count PubMed by nct_id
GET	/count/pmids/geneproteins	Yes	Count PubMed by gene_protein
GET	/organizations/{name}	Yes	Search organizations (wildcard *, ?, max 500)
GET	/organizations/assets/{org_id}	Yes	Organization assets (drugs, trials, max 500)
GET	/list/organization/aliases/{name}	Yes	DEPRECATED: Org alias search

Method	Path	Auth	Description
GET	/agent/api/health	No	Agent health check
GET	/agent/api/	No	Agent welcome
GET	/agent/api/auth/whoami	Yes	Agent user identity
POST	/agent/api/query	Yes	Synchronous agent query
POST	/agent/api/query/stream	Yes	Streaming agent query (SSE)

#	Tool	Class	API Endpoint Called	Purpose
1	fetch_identity()	EugeneIdentityTools	GET /auth/whoami	Validate user identity
2	fetch_by_label(label, page, page_size)	EugeneFetchTools	GET /labels/{label}	Browse drugs/diseases
3	fetch_similar(label, values)	EugeneFetchTools	POST /similarity/{label}	Find similar entities
4	lookup_node_by_value(value, fuzzy)	EugeneNodeTools	GET /node/find/{value}	Resolve name to ID
5	fetch_node_details(ids)	EugeneNodeTools	POST /node/details	Get node attributes
6	fetch_drug_aliases(drug_name)	EugeneDrugTools	GET /drugs/aliases/{name}	Drug synonym lookup
7	fetch_facts(node_id)	EugeneFactTools	GET /graph/facts/start/{id}	Paginated fact extraction (50 pages max)
8	fetch_node_relationships(node_id, n_hop)	EugeneGraphTools	GET /graph/relationship/start/{id}	N-hop traversal
9	fetch_paths(start, end, n_hop)	EugeneGraphTools	GET /graph/path/start/{s}/end/{e}	Path discovery
10	has_reachable_path(start, end, n_hop)	EugeneGraphTools	GET /graph/reachability/start/{s}/end/{e}	Reachability check
11	find_organization_names(pattern)	EugeneOrganizationTools	GET /organizations/{pattern}	Org wildcard search (10 pages)
12	find_organization_assets(org_id)	EugeneOrganizationTools	GET /organizations/assets/{id}	Org asset discovery (10 pages)

Tool Category	Purpose	Status
PUBMED	Direct PubMed article search via API	Code stub exists (ToolRequestEnum.PUBMED)
CLINICAL_TRIALS	ClinicalTrials.gov API integration	Code stub exists
CHEMBL	ChEMBL chemical database queries	Code stub exists
BIORXIV	BioRxiv preprint search	Code stub exists

Layer	Mechanism	Algorithm	Provider	Details
1. Enterprise SSO	Microsoft Entra ID	RS256 (asymmetric)	Microsoft	OIDC flow, JWKS validation with 8h cache, corporate identity
2. Platform Token	Eugene JWT	HS256 (symmetric)	Eugene Backend	24h TTL, claims: tid/sub/upn/roles, shared secret signing
3. Agent Allowlist	UPN Allowlist	N/A	Agent WS	Pipe-separated list of authorized user emails

Role	Scope	Assigned To	Status
user.public.read	Read public biomedical data	All authenticated users (default)	Active
user.confidential.read	Read proprietary CSL data	Named CSL users only	Defined, not yet enforced at query level

Control	Severity	Status	Detail
Cypher Injection	HIGH	OPEN	one-hop adapter uses string formatting instead of parameterized queries
API Rate Limiting	MEDIUM	OPEN	No rate limiting middleware on any endpoint
SAST/DAST Scanning	MEDIUM	OPEN	Checkov defined but SAST/DAST commented out in CI pipeline
TLS Certificates	LOW	ACTIVE	Self-signed certs on Neo4j and ALB; manual renewal
SSL Verification	MEDIUM	BY DESIGN	MCP and Agent HTTP clients use verify=False (internal network)
Static Role Mapping	LOW	OPEN	Roles hard-coded in roles.py; should migrate to DB or IdP
Input Validation	LOW	ACTIVE	Pydantic models + regex sanitization on org search
Secrets Management	LOW	ACTIVE	AWS Secrets Manager for production; .env for local dev

Provider	Default Model	Alternatives	Max Tokens	Temperature	Top-P
Anthropic	claude-sonnet-4-2025-0514	claude-haiku-4-5-2025-1001, claude-opus-4-20250514	16,384	0.7	N/A
OpenAI	gpt-4.1-mini	gpt-4.1, gpt-4o, gpt-4o-mini, gpt-3.5-turbo	4,096	0.7	1.0

#	Suggestion Text
1	List drug aliases for Adderall
2	Generate a table of drugs, diseases and clinical trial researched by Biogen Inc
3	What assets does Eugene know about companies with names like biogen
4	Does Mycophenolate mofetil have any relationships to PTRH2? If so explain
5	What recent pubmed studies mention ABL1?
6	What organizations are related to pmid 41402159?
7	What facts does eugene know about Sickel cell anemia?

Stage	Job(s)	Trigger	Actions
echo	debug	All branches	Log commit info
scan	checkov-scan	MR to qa/main	Checkov IaC scan (fail on HIGH/CRITICAL)
docker	build-push	qa, main, feature/*	Build Docker images, push to ECR (010928221940)
terraform-ec2	validate, plan	qa, main	Terraform init/validate/plan EC2 infra
apply-ec2	apply	qa, main (manual)	Terraform apply EC2 (manual approval gate)
terraform-services	validate, plan	qa, main	Terraform init/validate/plan ECS services
apply-services	apply	qa, main (manual)	Terraform apply ECS (manual approval gate)

Image	Dockerfile (Local)	Dockerfile (Prod)	Base	CMD
eugene_ws	docker/eugene_ws/	containers/eugene_ws/	python:3.13-slim	uvicorn eugene_ws:app --host 0.0.0.0 --port 8000
eugene_mcp	docker/eugene_mcp/	difflabs/containers/eugene_mcp/	python:3.13-slim	python src/eugene_mcp.py --host 0.0.0.0 --port 8000
eugene_agent_ws	docker/eugene_agent_ws/	difflabs/containers/eugene_agent_ws/	python:3.13-slim	uvicorn eugene_chat_ws:app --host 0.0.0.0 --port 8000
eugene_agent_ui	docker/eugene_agent_ui/	difflabs/containers/eugene_agent_ui/	python:3.13-slim	streamlit run src/eugene_agent_ui.py --server.port=8501
eugene_neo4j_ce	N/A (uses image)	containers/eugene_neo4j_ce/	neo4j:5.26.9-community	Sleep loop (manual start)

Module	Source	Resources Provisioned
eugene-ec2	infrastructure/modules/eugene-ec2/	EC2 instance (r7i.xlarge), EBS volumes, Security Groups, IAM Instance Profile
eugene-services	infrastructure/modules/eugene-services/	ECS Cluster, Task Definitions, ALB Target Groups, CloudWatch Logs

Module	Source	Resources Provisioned
eugene-containers (difflabs)	difflabs/iac/eugene-containers/	ECR repositories for all service images
eugene-services (difflabs)	difflabs/iac/eugene-services/	ECS services, load balancer targets

Parameter	Local Dev	difflabs (Dev)	AIA (Staging/Prod)
ENVIRONMENT	local	production	production
Neo4j URI	bolt://localhost:17687	bolt+ssc://10.88.203.244:7687	neo4j+ssc://<ec2>:7687
Neo4j Auth	neo4j/eugene_local_2024	AWS Secrets Manager	AWS Secrets Manager
Core API URL	http://localhost:18000	https://internal-eugene-search-alb-*.us-east-1.elb.amazonaws.com	https://eugene-staging.ai.cslg1.cslg.net
MCP Server URL	http://localhost:18443	https://<alb>:8443	https://<alb>:8443
Agent API URL	http://localhost:18001	https://<alb>/agent/api	https://<alb>/agent/api
Chat UI URL	http://localhost:18501	https://<alb>:8501	https://<alb>:8501
Auth Mode	Bypass (local token)	Entra ID OAuth	Entra ID OAuth
LLM Provider	Configurable	Configurable	Configurable
TLS	Disabled	Self-signed	Self-signed
Terraform Backend	N/A	S3 (us-east-1)	S3 (eu-central-1)
AWS Account	N/A	087084717211	010928221940

Service	Endpoint	Response	Docker Interval
neo4j	http://localhost:7474	HTML page (wget spider)	10s (10 retries, 30s startup)
eugene_ws	http://localhost:8000/health	{"status": "OK"}	10s (5 retries, 15s startup)
eugene_mcp	http://localhost:8000/health	{"status": "OK", "service": "eugene mcp service"}	On demand
eugene_agent_ws	http://localhost:8000/health	{"status": "OK", "service": "eugene agentic webservice"}	On demand

Canary	Endpoint	Validation
HealthEndpointCanary	GET /health	HTTP 200 OK response
StatsEndpointCanary	GET /stats	Valid node/relationship counts returned
LabelCountsEndpointCanary	GET /count/{label}	Non-zero counts for known labels
NodeFindEndpointCanary	GET /node/find/{value}	Node ID returned for known entity
NodeDetailsEndpointCanary	POST /node/details	Node attributes returned for known IDs

Setting	Value
Framework	pytest >= 7.0
Parallelism	pytest-xdist (-n 2 workers)
Coverage	pytest-cov (src/ directory, LCOV + terminal)
Markers	serial (no parallelism), integration (integration tests)
Environment	Loads .env via pytest-dotenv
Formatter	Black (line-length 88)
Linters	Ruff, mypy
Coverage Omit	*/*_test.py, */confest.py

Test Type	Scope	Priority
E2E Integration	UI -> Agent WS -> MCP -> Core API -> Neo4j	HIGH
JWT Edge Cases	Expired, invalid signature, missing claims, wrong audience	HIGH
Load Testing	Concurrent streaming requests, large result pagination	MEDIUM
LLM Failover	Anthropic unavailable -> OpenAI fallback	MEDIUM
MCP Unavailability	Agent behavior when MCP server is down	MEDIUM
Security	Cypher injection attempts, unauthorized access	HIGH

Source	Data Type	Ingestion Method	Status	Record Count
DrugBank	Drugs, aliases, interactions	CSV bulk import (neo4j-admin)	LOADED	4,600 drugs
Disease Ontology	Diseases, MONDO mappings	CSV bulk import	LOADED	17,100 diseases
Protein Databases	Genes, proteins, pathways	CSV bulk import	LOADED	27,700 genes
ClinicalTrials.gov	Trials, sponsors, conditions	CSV + Cypher	SCHEMA ONLY	49,300 nodes (empty)
USPTO	Patents, applications, claims	API download + Cypher	NOT INGESTED	0
PubMed/PMC	Articles, affiliations	API download + APOC	NOT INGESTED	0
CSL Internal	TPP, org data	Checkpoint files	PARTIAL	Various

#	Risk	Severity	Likelihood	Mitigation	Status
1	Cypher injection in one-hop adapter	HIGH	MEDIUM	Parameterize all Cypher queries using \$variables	OPEN
2	No API rate limiting	MEDIUM	MEDIUM	Add FastAPI middleware (slowapi)	OPEN
3	SAST/DAST scanning disabled	MEDIUM	HIGH	Re-enable in CI pipeline	OPEN
4	Self-signed TLS certificates	LOW	HIGH	Migrate to AWS ACM	OPEN
5	Incomplete data (patents, PubMed)	MEDIUM	HIGH	Execute pending ingestion pipelines	IN PROGRESS
6	AIA deployment broken	HIGH	HIGH	Fix ECS IAM permission issue	OPEN
7	Single MCP instance	MEDIUM	LOW	Implement S3SessionManager	PLANNED
8	Static role mapping	LOW	LOW	Migrate to DB-backed roles	OPEN
9	Neo4j CE limitations	MEDIUM	LOW	Evaluate Enterprise Edition for RBAC	DEFERRED
10	LLM provider outage	MEDIUM	LOW	Dual provider auto-fallback	PARTIAL
11	API key in docker.env	HIGH	LOW	.gitignore excludes; use Secrets Manager	MITIGATED
12	Python version variance	LOW	LOW	Standardize on Python 3.13	OPEN

#	Question	Context	Priority
1	When will patent and PubMed data ingestion be completed?	Code exists but data not loaded	HIGH
2	What is the plan to fix AIA ECS IAM permissions?	Production deployment blocked	HIGH
3	Should qa branch be merged into main?	Branches are desynced	HIGH
4	Should role-based filtering be at API or DB level?	Neo4j CE lacks native RBAC	MEDIUM
5	Should MCP scale horizontally (S3SessionManager)?	Currently single instance	MEDIUM
6	Should SAST/DAST be mandatory for all MRs?	Currently disabled in pipeline	MEDIUM
7	What is the TLS certificate renewal strategy?	Self-signed certs expire	MEDIUM
8	What additional MCP tools needed for MVP?	PubMed, ChEMBL stubs exist	MEDIUM
9	What is the target SLA for API response times?	No documented requirements	MEDIUM
10	Should clinical trial data be re-ingested or removed?	49.3K shell nodes exist	LOW

Variable	Service(s)	Required	Description
ENVIRONMENT	eugene_ws, agent_ws	Yes	local or production
NEO4J_URI	eugene_ws	Yes	Neo4j connection URI
NEO4J_USERNAME	eugene_ws, neo4j	Yes	Database username
NEO4J_PASSWORD	eugene_ws, neo4j	Yes	Database password
EUGENE_TENANT_ID	All backend	Yes	JWT issuer tenant ID
EUGENE_CLIENT_ID	All backend	Yes	JWT audience client ID
EUGENE_CLIENT_SECRET	All backend	Yes	JWT HS256 signing key
ENTRA_TENANT_ID	eugene_ws	Prod only	Azure AD tenant
ENTRA_CLIENT_ID	eugene_ws	Prod only	OAuth app registration ID
ENTRA_CLIENT_SECRET	eugene_ws	Prod only	OAuth client secret
ENTRA_AUTHORITY	eugene_ws	Prod only	Entra ID authority URL
ENTRA_SCOPE	eugene_ws	Prod only	OAuth scopes
REDIRECT_URI	eugene_ws	Prod only	OAuth callback URL
REDIRECT_PATH	eugene_ws	Prod only	OAuth callback path
EUGENE_API_BASE	eugene_mcp	Yes	Core API base URL
EUGENE_ISSUER	eugene_mcp	Yes	JWT issuer URL
EUGENE_AUDIENCE	eugene_mcp	Yes	JWT audience claim

Variable	Service(s)	Required	Description
EUGENE_MCP_SERVER_URL	eugene_agent_ws	Yes	MCP server URL
EUGENE_AGENT_ALLOWLIST	eugene_agent_ws	Yes	Pipe-separated user UPNs
EUGENE_AGENT_API_URL	eugene_agent_ui	No	Agent API base URL
LLM_PROVIDER	eugene_agent_ws	No	anthropic or openai
ANTHROPIC_API_KEY	eugene_agent_ws	Cond.	Anthropic API key
ANTHROPIC_MODEL_ID	eugene_agent_ws	No	Anthropic model (default: claude-sonnet-4-20250514)
OPENAI_API_KEY	eugene_agent_ws	Cond.	OpenAI API key
OPENAI_MODEL_ID	eugene_agent_ws	No	OpenAI model (default: gpt-4.1-mini)
HF_HUB_CACHE	eugene_ws	No	HuggingFace model cache directory
S3_STAGING_BUCKET	Production containers	Prod only	S3 bucket for .env and models

Version	Key Features
v1.0	Initial release: Neo4j graph database, foundational drug-disease-gene data
v1.1	Added patent search, PubMed search, drug alias endpoints
v1.2	Organization search and assets, graph facts endpoint
v2.0	MCP server with 12 tools, agent backend with Strands framework
v2.1	Streamlit chat UI, dual LLM support (Anthropic + OpenAI)
v2.2	Docker Compose local dev stack, CI/CD pipeline, Terraform IaC