## Lab 1: OCI Generative AI Medical Reasoning Data Pipeline

**Industry: Healthcare & Medical Education**

Story: Imagine you're part of a medical AI development team at a leading healthcare institution. Every day, medical professionals face complex diagnostic scenarios requiring deep reasoning, step-by-step analysis, and evidence-based decision making. The challenge? Processing vast amounts of medical case studies, diagnostic reasoning chains, and clinical knowledge to create an AI system that can understand and replicate expert-level medical thinking. This lab will show you how to use OCI Generative AI with OpenSearch to create a powerful medical reasoning pipeline that ingests, processes, and indexes medical case studies for advanced reasoning and training applications.

**Use Case:**

We will use OCI Data Science Notebooks, the OCI GenAI Embeddings Service, and OpenSearch to:

- Ingest and preprocess medical reasoning datasets containing complex clinical scenarios
- Convert medical questions, reasoning chains, and responses into high-quality vector embeddings
- Store them in an OpenSearch vector database for semantic retrieval and training data preparation
- Enable advanced medical reasoning model development and fine-tuning

**Dataset:**

The FreedomIntelligence/medical-o1-reasoning-SFT dataset from Hugging Face, which contains:

- Complex medical questions requiring multi-step reasoning
- Detailed reasoning chains (Complex_CoT) showing expert-level thinking processes
- Evidence-based medical responses with clinical accuracy
- Cases covering cardiology, neurology, pediatrics, pathology, and other specialties

**Workshop Steps (15 minutes):**

**Step 1: Ingest and Preprocess Medical Reasoning Data (5 mins)**

- Open OCI Data Science Notebook: Navigate to OCI Console → Analytics & AI → Data Science → Open Notebook Session already created
- Load Medical Dataset: Connect to Hugging Face and download the medical-o1-reasoning-SFT dataset

- Process Medical Case Structure: Implement Python code to parse the JSON format containing Question, Complex_CoT (reasoning chain), and Response fields
- Clean and Structure Medical Content: Apply preprocessing to extract medical entities, normalize terminology, and chunk reasoning chains for optimal embedding
- Validate Medical Data Quality: Ensure all cases contain complete reasoning chains and clinically accurate responses

**Step 2: Generate Medical Knowledge Embeddings (5 mins)**

- Connect to OCI GenAI Embeddings Service: Use the SDK or API to authenticate
- Create Specialized Medical Embeddings: Convert each medical question, reasoning step, and response into high-dimensional vectors optimized for medical semantics
- Process Complex Reasoning Chains: Embed individual reasoning steps to capture the logical flow of medical thinking
- Validate Medical Semantic Accuracy: Inspect embeddings to ensure medical terminology, diagnoses, and treatment concepts are properly captured

**Step 3: Store in OpenSearch for Medical Knowledge Retrieval (5 mins)**

- Deploy OpenSearch Cluster: Ensure your cluster is configured for medical data
- Create Medical Vector Index: Define mapping for vector fields, medical metadata (specialty, case type, difficulty), and diagnostic references
- Load Medical Embeddings: Push vectors + medical metadata into the specialized index
- Test Medical Queries: Try clinical scenarios like 'differential diagnosis for chest pain with ECG changes' and 'management of pediatric seizures' to validate retrieval accuracy

Conclusion & Value: In this lab, you've created an end-to-end AI-powered medical reasoning pipeline. You can now ingest complex medical cases with their reasoning chains, embed them for semantic understanding, and store them for instant retrieval. This foundation enables the development of medical AI systems that can replicate expert-level clinical reasoning and decision-making processes.

## Lab 2: OCI Generative AI Medical Reasoning Assistant with Advanced RAG

**Industry: Healthcare & Medical Education**

Story: Now that your medical knowledge base is operational, imagine being a medical educator or clinician who needs to train AI systems that can think through complex medical cases like experienced physicians. Instead of relying on simple question-answer pairs, you want an AI that can demonstrate the complete reasoning process—from initial symptoms through differential diagnosis to final treatment recommendations. In this lab, you'll use OCI Generative AI Agents with Retrieval-Augmented Generation (RAG) to create an advanced medical reasoning assistant that mirrors expert clinical thinking.

**Use Case:**

We will use the OpenSearch medical reasoning index from Lab 1 as the knowledge base for a conversational GenAI agent that:

- Demonstrates step-by-step medical reasoning for complex clinical scenarios
- Provides evidence-based diagnostic and treatment recommendations
- Shows the complete thought process from symptoms to conclusions
- Enables medical education and clinical decision support

**Dataset:**

The vectorized medical reasoning dataset from Lab 1 (Questions, Complex_CoT reasoning chains, and expert Responses from the medical-o1-reasoning-SFT dataset).

**Workshop Steps (10 minutes):**

**Step 1: Create the Medical Knowledge Base (3 mins)**

- Navigate to OCI Generative AI Agent Console
- Create Medical Knowledge Base: Point it to your OpenSearch medical reasoning index
- Configure Medical RAG Parameters: Define prompt templates for clinical reasoning, set retrieval parameters for medical case similarity, and format citations for medical evidence

**Step 2: Build and Deploy the Medical Reasoning Agent (5 mins)**

- Create Medical Agent: Give it a name (e.g., 'Clinical Reasoning Assistant')
- Enable Medical Reasoning Chains: Configure the agent to show step-by-step thinking processes like expert physicians
- Add Medical Tools: Connect to RAG search for case retrieval and reasoning chain analysis
- Configure Clinical Prompting: Set up prompts that encourage differential diagnosis, evidence-based reasoning, and systematic clinical thinking
- Test Medical Scenarios: Try complex prompts like 'A 45-year-old presents with chest pain and shortness of breath. Walk me through the complete diagnostic reasoning process including differential diagnosis, key investigations, and management approach.'

**Step 3: Validate Medical Reasoning Quality (2 mins)**

- Ensure responses demonstrate complete reasoning chains from symptoms to diagnosis
- Verify clinical accuracy and evidence-based recommendations
- Test ability to handle follow-up questions like 'What if the patient also had diabetes?' or 'How would you modify treatment for a pediatric patient?'
- Confirm the system shows uncertainty appropriately and recommends further evaluation when needed

Conclusion & Value: In this lab, you've transformed your medical reasoning dataset into a live, conversational AI medical reasoning assistant. With advanced RAG capabilities, the assistant demonstrates expert-level clinical thinking, provides step-by-step diagnostic reasoning, and serves as a powerful tool for medical education and clinical decision support. This system can help train the next generation of medical AI while supporting current clinical practice with evidence-based reasoning.

Key Success factors:

✅ **Use specific clinical terminology** (dataset is medical-focused)
✅ **Request evidence-based reasoning** (matches the Chain-of-Thought structure)
✅ **Ask for concise responses** (aligns with response format)
✅ **Focus on actionable clinical decisions**
✅ **Include 'why' or 'explain'** for deeper reasoning

Example prompts

- **Diagnostic reasoning:**
*"Given a patient with sudden neurological symptoms after long-distance travel and evidence of leg swelling, what cardiac abnormality could explain these findings?"*

- **Injury localization:**
*"A patient sustains a penetrating chest wound near the 8th rib in the midaxillary line. Which thoracic structure is most likely injured?"*

- **Test interpretation:**
*"In a woman with stress urinary incontinence confirmed by Q-tip test, what would cystometry most likely reveal regarding residual volume and detrusor contractions?"*

- **Differential diagnosis with history:**
*"A 45-year-old man with past alcohol use presents with sudden dysarthria, shuffling gait, and intention tremor. What is the most likely diagnosis?"*

- **Pathological findings:**
*"A 45-year-old man presents with parkinsonism symptoms, hallucinations, and memory issues. What histological brain finding would most likely be observed?"*