Keeping up with
AIverything

The adventures of a 50-something trying to keep up with AI

AI Foundations – Oracle OCI – A Short Course

ation

the content for the

tions Associate

ific AI knowledge and

finish off with a summary of ethical considerations when dealing with AI.

Here is the link to the certification if you are interested (you will need an Oracle account): Oracle Cloud Infrastructure AI Foundations Associate 2025 certification

Let’s begin.

Module 5: AI on Oracle Cloud Infrastructure (OCI)

Introduction

Having covered the theory behind AI, Machine Learning, and Deep Learning in the earlier article, this module brings those concepts into the real world. We will explore the comprehensive suite of AI tools, platforms, and services available on Oracle Cloud Infrastructure.

See this as a bridge from “what it is” to “how you do it” on OCI, covering the infrastructure that powers AI, the pre-built services that make AI accessible, and the advanced platforms for building custom solutions.

The Foundation: OCI AI Infrastructure

Powerful AI models require powerful hardware. OCI provides the high-performance, scalable infrastructure necessary for demanding AI workloads, from small-scale experiments to training massive foundational models.

Compute (GPUs): Graphics Processing Units (GPUs) are the workhorses of AI, providing the parallel processing power needed to train deep neural networks efficiently. At the time of writing, OCI offers a range of NVIDIA GPUs tailored to different needs:

- **For small to medium workloads:** The *NVIDIA A10* and *NVIDIA A100* are common choices for general AI training and inference.
- **For massive-scale and HPC workloads:** More powerful GPUs like the *NVIDIA H100* or *NVIDIA GB200* are used.
- **Networking (OCI Supercluster):** For the most complex AI tasks, like training large language models, OCI offers *Superclusters*. These are designed to deliver exceptional, large-scale performance by connecting thousands of GPUs with ultra-low-latency networking, allowing them to work together as a single, massive supercomputer.

Pre-trained AI Services: AI with No ML Expertise Required

OCI offers a suite of pre-trained, API-driven services that allow developers to easily add AI capabilities to their applications without needing deep machine learning expertise.

OCI Vision: For Understanding Images and Documents

Core Function: Analyses visual data

Object Detection: Locating and labelling specific objects within an image (e.g., identifying vehicles and license plates in a security camera feed).
Image Classification: Assigning labels to an entire image based on its content.
Document Understanding: A specialized feature for working with documents. Key capabilities include:

- **Key-Value Extraction:** Automatically finding and extracting specific details from documents like receipts (e.g., merchant name, date, total amount).
- **Document Classification:** Identifying a document's type (e.g., invoice, receipt, or resume) based on its layout and keywords.

- **Optical Character Recognition (OCR):** Recognizing and extracting printed or handwritten text from a document.

OCI Speech: For Converting Audio to Text

Core Function: Provides highly accurate, automated speech-to-text transcription.

Batch Support: Process thousands of audio files efficiently in a single job, rather than one by one.
Text Normalization: Improves readability by converting numbers, dates, URLs and monetary amounts into standard formats.
Confidence Scoring: Provides a score for each transcribed word, allowing you to gauge the model's certainty, which especially important in environments where accuracy is critical such as legal or medical contexts.
SRT File Support: Generates output in the SRT format, which is the standard for creating closed captions for video content.
Profanity Filtering: Can *remove, mask, or tag* profane words in a transcription based on user preference.

OCI Language: For Understanding Unstructured Text

Core Function: Extracts insights and structure from text.

Text Classification: Automatically categorizing articles or documents into predefined topics e.g., "Politics", "Technology", "Sports".
Other key features include: Sentiment Analysis, Named Entity Recognition, Personal Identifiable Information Detection and Key Phrase Extraction. Note that this service is for *text analysis*, not *text generation.*

OCI Platforms for Custom AI Solutions

For data scientists who need to build, train, and deploy their own models, OCI provides end-to-end platforms.

OCI Data Science Service: The ML Workbench

Purpose: An integrated environment for the entire machine learning lifecycle.

Key Components:

- *Notebook Sessions*: An interactive coding environment (based on JupyterLab) for building and training models.
- *Model Catalog*: A central repository for storing, tracking, sharing, and managing machine learning models and their artifacts.
- *Model Deployment*: The capability to deploy trained models from the catalog as real-time HTTP endpoints for inference.

OCI Generative AI Service: The LLM Platform

Purpose: A platform for using and customizing large language models.

Key Features:

- *Access to Foundational Models*: Provides a choice of pre-trained models for different tasks (e.g., Chat models, Embedding models). It does not, however, have a specific category for “Translation models”.
- *Fine-tuning*: Allows you to adjust a pre-trained model’s parameters on your own data to improve its accuracy for a specific task using high-performance **dedicated AI clusters**.
- *Embedding Models*: These models are used to convert text into numerical vectors for tasks like *semantic search*, where you can find information based on meaning rather than keywords.

AI in the Oracle Database: Bringing Intelligence to Your Data

Oracle integrates AI capabilities directly into its flagship database, allowing you to perform AI tasks where your data already resides.

Oracle AI Vector Search

Purpose: Integrated into Oracle Database 23ai, *Vectors* enable *semantic search*. Instead of querying for exact keywords, you can query based on contextual meaning. For example, finding all documents “about high-risk investments”

even if they don't use that exact phrase.
Integration: Pre-trained models can be used by allowing them to be loaded directly into the database in standard formats like *ONNX* (Open Neural Network Exchange).

Think of ONNX (Open Neural Network Exchange) as a universal translator for AI models.

- **The Problem It Solves:** Different deep learning frameworks (like PyTorch, TensorFlow, etc.) save their trained models in their own unique formats which makes it difficult to train a model in one framework and use it in a different tool or on a different platform.
- **The Solution It Provides:** ONNX is an open-source, common format that acts as a bridge. A developer can train their model in their preferred framework, convert it to the ONNX format, and then confidently load and use that model in any other system that supports ONNX, like Oracle Database 23ai.
The key benefit is interoperability. It gives developers the freedom to choose the best tool for each part of the job without being locked into a single ecosystem.

Oracle Select AI

Purpose: *Select AI* transforms how users interact with the Autonomous Database by functioning as an intelligent *natural language* interface. Its core function is to translate user questions, asked in plain English, into executable SQL code, effectively *turning the database into a conversational partner*.

How It Works: From English to SQL

The process is more than a simple translation; it's a full workflow that uses the power of a Large Language Model (LLM).

1. **The User Asks a Question:** A user, who may have no knowledge of SQL, poses a question in natural language. For example, *"Show me the top 5 performing products in the UK for the last quarter."*
2. **Select AI Gathers Context:** Select AI doesn't just send the raw question to an LLM. It first *gathers relevant database schema metadata*: information about the

tables, columns, their data types, and relationships (e.g., it knows the SALES table joins with the PRODUCTS table).

3. **It Connects to an LLM:** Next, the user's question *along with the relevant schema context* is *sent to a powerful LLM*. The context is vital because it gives the LLM the map it needs to understand what data is available and how it's structured.
4. **The LLM Generates the SQL Query:** Based on the user's intent and its understanding of the database schema, the LLM formulates the precise SQL query required to answer the question.
5. **Execution and Results:** Select AI takes the generated SQL, runs it against the database, and returns the final answer to the user in a readable format.

Key Benefits

- **Democratizes Data Access:** The primary advantage is that it empowers non-technical users, such as business analysts, managers, and executives, to get insights directly from complex data without needing to learn SQL or rely on a developer. It breaks down the barrier between business questions and data-driven answers.
- **Increases Productivity:** Even for experienced developers, writing complex queries with multiple joins and filters takes time. Select AI can generate an accurate first draft of a query in seconds, which the developer can then use or refine, significantly speeding up development and analysis cycles.
- **Enhances User Interaction:** It makes interacting with a database more intuitive and conversational, lowering the learning curve and encouraging more widespread data exploration within an organization.

Module 6: Responsible and Ethical AI

Introduction

Having explored the powerful capabilities of AI, from predictive machine learning to creative generative models, we now turn to one of the most critical aspects of the field: *building AI that is trustworthy*.

Technology alone is not enough; for AI to be successfully integrated into society, it must be *developed and used responsibly*. This final module covers the essential principles and practices that ensure AI systems are lawful, ethical, and robust.

The Three Pillars of Trustworthy AI

For an AI system to be considered trustworthy, it must be built on a foundation of three core guiding principles. These pillars ensure that AI operates safely, fairly, and in alignment with human values and legal frameworks.

1. Lawful AI

Principle: AI systems must comply with all applicable local, national, and international laws and regulations.
In Practice: This includes adhering to data privacy laws (like GDPR), respecting intellectual property rights, ensuring accessibility standards are met, and complying with industry-specific regulations (such as those in finance or healthcare). It's the baseline for any legitimate AI application.

2. Ethical AI

Principle: AI usage must go beyond legal compliance to ensure AI systems align with fundamental ethical principles and societal values. It is about doing what is right, not just what is legally permissible.
In Practice: There are several key considerations:

- *Respect for Human Autonomy*: AI should augment human capabilities, not replace human self-determination. Systems should be designed with human oversight in mind.
- *Prevention of Harm*: AI should be safe and not cause physical, psychological, or

financial harm, and should also include ensuring the system is secure from malicious attacks.

- *Fairness*: AI systems should treat all individuals and groups equitably and avoid creating or reinforcing unfair bias.

3. Robust AI

Principle: Also critical is the technical reliability and resilience of the AI system. A trustworthy system must work correctly, consistently, and securely.
In Practice: A robust AI system is one that is:

- *Accurate and Reliable*: It performs as expected and produces accurate results for its intended purpose.
- *Reproducible*: Its results can be reproduced under similar conditions, which is key for validation and testing.
- *Resilient*: It can withstand unexpected conditions or adversarial attacks designed to manipulate its behaviour.

AI Ethics in Practice: From Principle to Requirement

High-level principles are essential, but they must be translated into actionable requirements for the teams building and deploying AI. Understanding how a principle connects to a practical requirement is key and a prime example is the relationship between *Explicability* and *Transparency*.

The Principle: Explicability

Definition: Explicability (often used interchangeably with "Explainability") is the principle that it should be possible to understand and explain *why an AI model made a particular decision*. For complex "black box" models like deep neural networks, this can be challenging but it is crucial for building trust, debugging errors, and ensuring fairness. We need to be able to look inside the box.

The Requirement: Transparency

Definition: Transparency is the practical requirement of being open and clear with users and stakeholders about how an AI system works, which includes communicating what data it uses, what its capabilities and limitations are, and who is accountable for it.

The Connection:

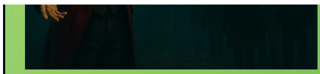
You cannot be transparent about a process you cannot explain. Therefore, Explicability and Transparency are tightly linked. By being able to explain a model's decision-making process, an organization can then be transparent with its users about why a certain outcome occurred (e.g., why a loan application was denied).

The "principle-to-requirement" link is fundamental to Responsible AI. The principle of *Fairness* leads to the requirement of *Bias Auditing*, and the principle of *Prevention of Harm* leads to the requirement of *Rigorous Safety Testing*. Understanding this framework is key to putting AI ethics into practice.

Conclusion

That's the basics of AI through an Oracle Cloud Infrastructure lens. Given the information in this article, you should have a better understanding of what you are reading in other stories about Oracle AI.

If you intend to carry on and sit the **Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate** exam, do read through the first article in this series which covers the fundamentals of AI terminology and practical use cases.



LET’S DO THIS!

We don’t spam. Read our privacy policy for more info.

← Previous Post

Next Post →

About

Terms & Conditions

Privacy Policy

Disclaimer

Contact