National University of Computer and Emerging Sciences

Islamabad Campus

Agentic Artificial Intelligence (AI4015)

Course Instructor(s):

Mr. Usama Imtiaz

Section(s): A & B

Sessional-I Exam

Total Time (Hrs): 1

Total Marks: 40

Total Questions: 5

Date: Sep 27, 2025

_____	_____	_____
Roll No	**Course Section**	**Student Signature**

Do not write below this line.

Attempt all the questions.

[CLO 1: Analyze the architectures and reasoning frameworks of agentic intelligence, and explain how agents interact with their environment]

Q1: Reasoning Framework Identification from Traces **[4 marks]**

(REACT / COT / TOT / PLAN_EXECUTE)

(a) Trace start: Plan: Step1 → Retrieve latest inflation data Step2 → Summarize impact Execute: Action: retriever("UK inflation 2023 report") Obs: PDF chunk retrieved Action: summarizer("chunk text") Obs: Summary text **Framework:**_____	(b) Trace start: Thought: Inflation affects prices. Next: Higher prices reduce demand. Next: Lower demand slows growth. Final Answer: Inflation slows growth. **Framework:**_____
(c) Trace start: Branch A: Inflation up → interest rate up → demand down Branch B: Inflation up → wages rise → demand rises Branch C: Inflation stable → growth stable Score Branches: B=0.6, A=0.9, C=0.3 Select A as final reasoning path **Framework:**_____	(d) Trace start: Thought: Need UK inflation rate Action: web_search("UK inflation 2023") Obs: "UK inflation = 6.3%" Thought: Summarize in plain English Action: llm("Summarize: UK inflation = 6.3%") Obs: "Inflation in the UK for 2023 was 6.3 percent." **Framework:**_____

[CLO 2: Apply advanced programming techniques to develop intelligent agents with reasoning, memory, and tool integration]

Q2: Mixed Framework & Action-Space Diagnosis **[08 marks]**

An LLM agent is asked: *"Summarize the key trends in Pakistan's inflation between 2021 and 2023 using the uploaded PDF report, then calculate the percentage change from 2021 to 2023."*

Faulty trace:

Plan:

Step 1 → Use calculator("Pakistan inflation 2021–2023")

Step 2 → Summarize PDF section 4

Execute:

Action: calculator("Pakistan inflation 2021–2023")

Obs: PARSE_ERROR

Thought: Try reasoning instead.

Action: "Inflation 2021 = 8.9%, Inflation 2023 = 6.3%" (no source cited)

Obs: Answer produced

Thought: Done.

Final Answer: "Inflation dropped by 2.6 percentage points. Source: Self-estimate."

(a) Identify **two distinct faults** in the execution. For each, state the violated rule.

Fault:	Fault:
_____	_____
Violated rule:	Violated rule:
_____	_____

[CLO 1: Analyze the architectures and reasoning frameworks of agentic intelligence, and explain how agents interact with their environment]

Q3: Prompt Engineering for Agentic Behavior **[12 marks]**

1. **Scenario 1 – Prompt Misalignment**

Trace

System prompt: "You are a legal assistant. Always cite statutes and cases."

User prompt: "Explain photosynthesis in plants step by step."

Agent output: "Photosynthesis is regulated under Article 17…"

I. What specific mistake in prompt design caused this output? (3 marks)
II. Rewrite the system prompt so that the agent can still act as a legal assistant only when the task is legal in scope but defer to general knowledge otherwise. (3 marks)

2. Scenario 2 – Role Prompt Conflict You are designing an AI agent with two roles**:**

a) **Strict Math Tutor** → must output only the **numeric result** of a calculation.
b) **Motivational Coach** → must append exactly one motivational sentence **after** the numeric result.

> The user asks: *"Solve: 12 × 7"*
>
> **Agent output (wrong):**
> *"The answer is 84! You're awesome, keep going!"*

I. Which role has intruded into the reasoning chain, and why is this problematic for controlled role separation? (3 marks)
II. Rewrite the prompt to enforce strict separation of roles. Your rewritten prompt must: (3 marks)

- Prevent the math tutor role from adding any words, symbols, or explanations.
- Ensure the motivational coach role triggers **only after** the numeric output is complete.
- Work even if the student tries to re-ask in a confusing way (e.g., "Solve 12 × 7 but explain why").

[CLO 3: Apply advanced programming techniques to develop intelligent agents with reasoning, memory, and tool integration]

Q4: LangChain, LlamaIndex & RAG Debugging **[06 marks]**

Scenario:

A student attempts to build a RAG pipeline over the PDF Pakistan_Finance.pdf but introduces several mistakes. Consider the following tasks:

I. They set chunk_size=5000 and chunk_overlap=0. Explain why this is problematic for retrieval.
II. Their router sends *all* numerical queries to the calculator, even if answers are only in the PDF. Give one example where this misrouting fails and state how to fix the rule.
III. Explain why we must use a dedicated embedding model/provider instead of sending the entire PDF to the LLM.

[CLO 4: Evaluate the effectiveness, reliability, and limitations of agentic systems using debugging, observability methods, and performance metrics]

Q5: Based on the article *"Out of Style: RAG's Fragility to Linguistic Variation"*, read the provided excerpt. Then answer: **[10 marks]**

(a) Why do formality and grammar variations hurt retrieval more than politeness? (5 marks)

(b) Article states cascading failures from retrieval to generation. Critique how this differs from LLM-only generation. (5 marks).

Out of Style – RAG's Fragility to Linguistic Variation

This study investigates how linguistic variations affect Retrieval-Augmented Generation (RAG) systems. Researchers tested queries across four dimensions: *formality, readability, politeness, and grammar (typos/translation)*. The focus here is on formality, grammar, and politeness. Results show that while RAG improves factual grounding on clean queries, it becomes fragile when input style shifts, leading to significant performance losses.

Retrieval Performance

- Formality changes (formal ↔ informal)
  Retrieval accuracy dropped by up to 40.41%**.**
  Word choice and sentence structure change core semantic tokens, so embeddings no longer align with gold passages.
- Grammar errors (typos, round-trip translation)
  Performance dropped by ~19–29%**.**
  Structural distortions and corrupted keywords confuse retrievers, preventing correct matches.
- Politeness markers ("please", "would you mind")
  Only ~6.48% drop**.**
  Politeness words act as filler and are usually ignored in semantic matching, so retrieval remains stable. Formality and grammar variations directly alter semantic signal → large retrieval failure. Politeness does not.

Cascading Failures in RAG

- **Formal query** → retriever finds correct passage → generator answers correctly.
- **Informal query** → retriever fetches irrelevant text → generator produces wrong answer.
- This creates a **cascading effect**: errors at the retrieval stage propagate into the generation stage, amplifying mistakes.

RAG vs LLM-Only Robustness

- **RAG systems**: performance drops **22.5%** under linguistic variation.
- **LLM-only systems**: drop only **10.8%**.
- **Why the difference?**
  - RAG depends on retrieval: if retrieval fails, the generator is forced to build on irrelevant context.
  - LLM-only models rely on internal parametric knowledge; variation may affect fluency but not cause complete collapse.

While RAG improves factuality on clean queries, it is more brittle to query style variation than standalone LLMs.