EdSutra

Teaching guide

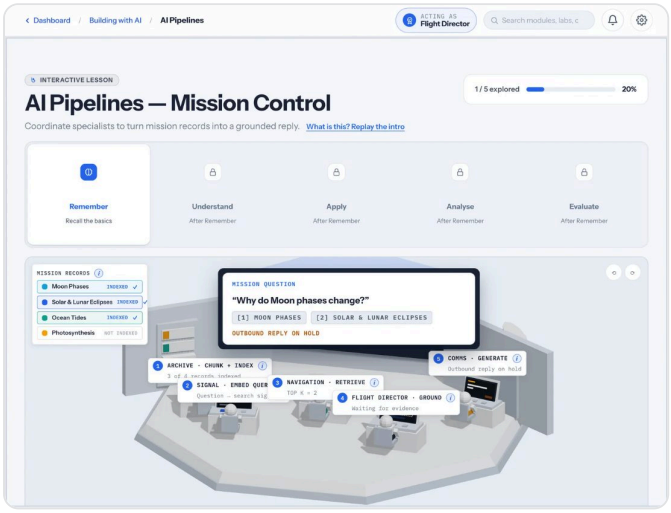
Module 06 · Building with AI · Unit 03 · Stage 3, Responsibility

AI Pipelines

Mission Control

Grades 9–12 | 50 minutes | Role: Flight Director

Grade 9: About age 14–15 · Year 10 in the UK · Class 9 in India



The big idea

An AI pipeline is a relay of specialists: index the records, turn the question into search terms, retrieve, ground, generate. Retrieval can bring in facts the model lacks, but it cannot force later stages to use them. When a reply goes wrong, trace it back to the first broken handoff and fix that stage.

The learner's job

Learners are the **Flight Director** at Kestrel Works · Mission Control. Every stage says it is fine. The thing coming out the end is not. Their task: find the stage where the reply first went wrong. Fix that one.

By the end, learners can say

- I can name what each pipeline specialist contributes: index, embed, retrieve, ground, generate
- I can predict which failures retrieval can fix — and which it cannot guarantee
- I can operate a RAG system: tune TOP K, inspect scores, and build a cited evidence brief
- I can trace a bad reply to the first broken stage, citing the artifact that proves it
- I can choose and defend a sequential, parallel, or conditional architecture under constraints

Before class

- One computer or tablet for each learner or pair, with EdSutra open (this Module is part of the full EdSutra studio).
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play the first two missions yourself, so you know which consoles learners click and what each prediction case shows.
- Write the five specialists on the board in order: Archive, Signal, Navigation, Flight Director, Comms. Learners add each one's job in Remember.
- Pairs work well: one learner operates, the Flight Director names the stage they suspect before anyone checks.
- Best after Build Your AI and Agentic AI.
- Units open in order, and Module 1 comes first. To start a class here, switch on Open every Unit in Settings → Teachers & parents.

Words to introduce

Pipeline: A chain of specialists, each doing one job and handing work on.

Retrieval: Searching stored records for the ones that best match a question.

TOP K: How many of the best-scoring records the search returns (1 to 3 here).

Grounding: Attaching the retrieved records to the brief, so every claim cites a source.

RAG: Retrieval-augmented generation: search first, then write the reply from what was found.

Fallback: A safe reply, like "I don't have enough evidence", when nothing relevant is found.

Architecture: How stages are arranged: one after another, side by side, or only when needed.

Running the lesson 50 minutes

Open 5 min

"Every stage says it is fine. The reply coming out the end is wrong. How would you find where it went wrong?"

Before anyone clicks, ask: if a chatbot can search the right documents before answering, will its answers always be right? Take a hands-up vote and record it. Come back to it in Understand, after the ignored bulletin.

Step	What learners do	Ask the class	Listen for
1. Remember 6 min	Click each console in Mission Control to meet the five specialists: Archive, Signal, Navigation, Flight Director and Comms. Then answer: what must the Flight Director give Comms?	• Which specialist finds the records, and which one writes the reply? • Why does Comms need the records, not just the question?	Comms writes from the question plus the retrieved records. That handoff is grounding, so every claim can point to a source.
2. Understand 8 min	Two cases. For each, predict whether retrieval will rescue the reply, then watch the replay: an outdated rock-sample answer, and a safety bulletin that is retrieved but never used.	• Why did retrieval fix the rock-sample answer? • The bulletin was found. Why was the reply still wrong? • Look at our vote. Were we right?	Retrieval helps when the fact is in the records. It cannot make later stages use what it found.
3. Apply 11 min	Pick a brief, choose which records are indexed, set TOP K, and run the search. Read every keyword score, attach only useful records, and generate a grounded reply.	• What changed when you set TOP K to 1, then 3? • What happened when you attached only the Eclipses record?	Scores are counted keyword matches. More records can help or distract, and with no relevant evidence the reply says so.
4. Analyze 10 min	Inspect three bad replies. For each, follow the record from Archive to Comms and pick the first broken stage, the evidence that proves it, a repair and the reason.	• Was the Auroras record ever in the Archive? • In the crater case, which stages worked, and which one broke?	The first break, not the last: wrong record at Navigation, dropped bulletin at the Flight Director, reversed claim at Comms.
5. Evaluate 10 min	For all three scenarios (routine check-ins, research reports, emergency support), pick sequential, parallel or conditional, simulate the messages, tag at least two trade-offs, keep the no-evidence fallback, and choose a matching defense.	• Why does running two models on routine check-ins waste compute? • In emergency support, which messages go to the expert, and why?	The best design spends extra compute only where it buys reliability. More branches are not automatically better.

Steps follow Bloom's revised taxonomy, in the order the Unit presents them.

Deepen the thinking

Critical thinking	In Analyze, the stage nearest the bad reply was not always the broken one. Where else might blaming the last step hide the real cause?
AI thinking	This search counts matching keywords. The question about glowing skies near the poles never says "aurora". How could a search find records by meaning instead?
Ethics and fairness	A school help bot answers from an outdated handbook and a learner misses a deadline. Who is responsible: the bot's builders, the handbook's owners, or the learner?
Connect to life	Think of a group project where each person did one part. How did a mistake early on travel through to the final result?

Misconceptions to listen for

"Retrieval stops AI from making things up."
In Understand, the bulletin is retrieved but never reaches the brief, and in Analyze, Comms reverses a claim it was given.

"More evidence is always better."
At TOP K 3, the search also returns Ocean Tides for a Moon phases question. Extra records can distract, so attach only what answers it.

"The stage nearest the bad reply is the broken one."
In the auroras case, Comms wrote the wrong reply, but the break came earlier: Navigation returned the wrong record. Fix the first break.

In real life the card learners see at the end of the Unit

- Some chatbots search documents first and show the sources they used: retrieval plus generation.
- A school or company help bot answering from its handbook gives outdated answers if the handbook is outdated.
- At a library help desk, finding the right book comes before the answer; if that step fails, everything after does.

Try it this week. Answer a question using only two textbook pages, noting which page backs each sentence. Did any sentence come from neither?

Watch out for. Showing sources doesn't guarantee a right answer. Retrieval can fetch the wrong record, and the earliest broken step is the one to fix.

Talk about it. Would you trust an answer more if it showed its sources? What would you still check?

Exit ticket

1. What does the Flight Director add to the brief before Comms writes?

2. Retrieval found the right record. Name one way the reply can still be wrong.

3. When is a conditional pipeline a better choice than a parallel one?

Going further

The ML Pipeline Builder Lab lets learners fill four stages, run them for a real score, and swap one block at a time to see which stage matters. For homework, use the In real life card's "Try it this week".