EdSutra

Teaching guide

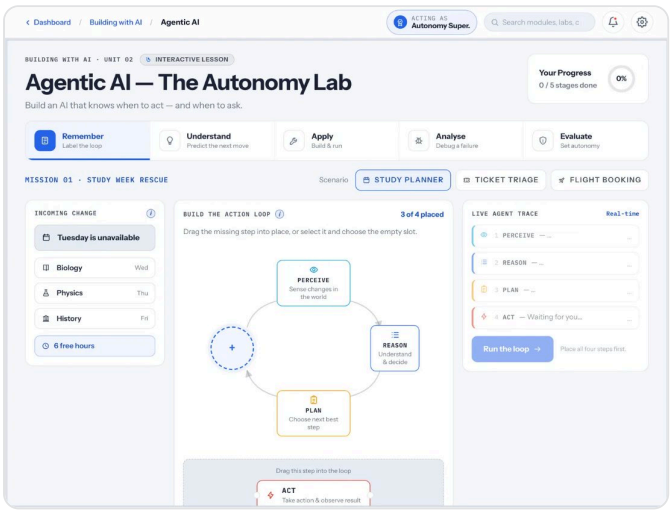
Module 06 · Building with AI · Unit 02 · Stage 3, Responsibility

Agentic AI

The Autonomy Lab

Grades 9–12 | 40 minutes | Role: **Autonomy Supervisor**

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



The big idea

An agent does not just answer: it runs a loop (perceive, reason, plan, act) until a rule says stop. Humans decide which tools it gets, when it must stop and when it must ask first. The right amount of autonomy depends on the stakes and how easily a mistake can be undone.

The learner's job

Learners are the **Autonomy Supervisor** at Kestrel Works · Autonomy Bay. This machine does not wait for permission between steps. Someone decides how far it gets. Their task: decide how much this system may do alone. Then watch it do it.

By the end, learners can say

- I can name the four parts of the agent action loop — Perceive, Reason, Plan, Act
- I can predict how a reactive agent and a planning agent respond when a plan fails
- I can construct a bounded agent with a scoped tool and an explicit stopping condition
- I can trace an agent failure to the part of the loop that broke, citing transcript evidence
- I can choose an autonomy level and guardrail for a deployment and defend the trade-off

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 steps yourself, so you know where learners will drag and what the two agents do.
- Draw the loop on the board: Perceive, Reason, Plan, Act, with an arrow back to Perceive. Leave space to add Stop in Apply.
- Pairs work well: one learner drives, the Autonomy Supervisor decides what the agent may do alone, and says why.
- Best after Build Your 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

Agent: An AI that keeps choosing its next step until a job is done.

Action loop: Perceive, Reason, Plan, Act, then check the result and go again.

Tool: Something the agent is allowed to use, such as a calendar.

Stopping condition: A clear rule that tells the agent when the job is done.

Least privilege: Give the agent only the tools this one job needs.

Guardrail: A rule people set so an agent's mistakes stay small.

Autonomy level: How much the agent does before a person checks: act, report, or ask first.

Running the lesson 40 minutes

Open 4 min

"This machine does not wait for permission between steps. Who decides how far it gets, and how would they decide?"

Before anyone clicks, ask: an AI planner moves your study blocks after Tuesday becomes busy. Everything now fits. How does it know it is finished? Write the guesses down and come back to them in Apply.

Step	What learners do	Ask the class	Listen for
1. Remember 5 min	Pick a scenario (study planner, ticket triage or flight booking). Drag the missing Act step into the loop, then press Run the loop and watch the live agent trace fill in.	• What did the agent perceive first? • Where does the result go after Act?	Each part has a job. The result feeds back to Perceive, and the loop repeats until the job is done.
2. Understand 7 min	For each case, predict how a reactive agent and a planning agent handle a broken step, then watch both lanes play out. Learners finish at least two of the three cases.	• What did the reactive agent do when Wednesday was full? • In the flight case, when did each agent check the budget?	A reactive agent has one move for one trigger. A planning agent holds a goal and can reroute.
3. Apply 8 min	Run the starter agent first: it plans the week, then has no rule to stop, and pauses at the app's safety cap. Attach the stopping condition and run it again.	• Why is the Refund tool locked for a study planner? • Look at our guesses from the start. How does the agent know it is finished?	Without a stop rule, a loop is stuck, not autonomous. The agent gets only the tools this job needs.
4. Analyze 8 min	Read three failure transcripts row by row. For each, pick the broken part of the loop, the evidence that proves it, and the guardrail that would have prevented it.	• In the order case, where did "arrives Friday" come from? • Which row shows the agent going beyond its task?	Infinite loop breaks at Stop, the invented order status at Perceive, the photo clean-up at Act. Each fix is a guardrail.
5. Evaluate 8 min	For at least two of three agents (refunds, coding, study planner), choose an autonomy level, pick and rank a guardrail, set the threshold, tag the evidence and pick a reason that backs the choice.	• Why might the study planner act alone while the refund agent asks first? • What does a guardrail cost, and what does it save?	Autonomy follows the stakes: real money and hard-to-undo actions need a person sooner. More than one answer can be defended.

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

Deepen the thinking

Critical thinking	The starter agent looked fine for several steps and still had no way to end. How would you test an agent before trusting it with a real job?
AI thinking	A reactive agent gets stuck when one step breaks; a planning agent can reroute. When might a simple reactive agent still be the better design?
Ethics and fairness	When an agent's mistake hurts someone, who is responsible: the agent, the person who set its autonomy, or the company that gave it the tool?
Connect to life	A robot vacuum senses, plans, acts and checks, again and again. What other machine or routine in your life runs a loop with a stop rule?

Misconceptions to listen for

"More autonomy is always better."
In Evaluate, the possible damage spreads further as autonomy rises. A study planner can act alone; a refund agent may need a person to confirm.

"A demo that works means the agent is ready."
In Apply, the starter agent plans the week perfectly, then has no way to end. Test what happens when things go wrong, and how it stops.

"Guardrails just slow the agent down."
In Analyze, every failure's fix is a guardrail, and in Apply the run cannot finish without a stop rule. Guardrails make autonomy safe to use.

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

- Some AI assistants can now book, buy or send things, looping through steps until a stopping rule says done.
- A robot vacuum senses a room, plans a path, acts, checks the result and repeats: an agent loop.
- An email helper that suggests a reply and one that sends it for you have different autonomy.

Try it this week. Write a step-by-step loop for a chore, like packing a school bag. Add a stop rule and one step where a person should check.

Watch out for. More autonomy isn't always better. The higher the stakes, and the harder a mistake is to undo, the sooner an agent should ask a person.

Talk about it. What's one job you'd let an AI agent do alone, and one it should always ask about?

Exit ticket

1. Name the four parts of the agent loop, in order.

2. Why does an agent need a stopping condition?

3. What does least privilege mean? Give an example from the Unit.

Going further

The Open Project Canvas Lab lets learners build an agent loop from simple rules and watch it run. For homework, use the In real life card's "Try it this week": a chore loop with a stop rule and a human check.