AGENT AI SURVEY

# Agent AI Is Not Just a Chatbot

A classroom reading of Agent AI: Surveying the Horizons of Multimodal Interaction

## KEY TEACHING POINTS

- Perception: what can the system observe?
- Action: what can it change?
- Feedback: what does it learn from next?

### Perception

Reads language, images, audio, video, or environment state.

### Action

Changes a physical or digital environment.

### Feedback

Observes the result and adjusts the next step.

**Agent AI participates in a task loop; it does more than return advice.**

2024 PAPER FRAMING

# The Paper Gives a Map, Not a Finished Machine

Use the January 2024 survey as conceptual vocabulary—not as a 2026 catalog of current systems.

## KEY TEACHING POINTS

- Moves beyond the beginner formula “LLM + tools + memory.”
- Centers multimodal perception and grounded environments.
- Connects action prediction, feedback, evaluation, and risk.

### Foundation model

A building block, not the entire agent.

### Multimodal perception

Language plus sensory and contextual signals.

### Grounded action

Behavior occurs inside a physical or virtual environment.

### Evaluation in context

Measure outcomes, feedback, and risk—not only text quality.

**Historical framing: durable concepts, dated examples.**

CONCEPT CORRECTION

# Multimodal Is Not Automatically Agentic

The difference is the task loop—not simply the number of media types a model can read.

## KEY TEACHING POINTS

- A model can understand images or video without choosing an action.
- An agent uses perception, state, goals, action, and feedback.
- A caption is useful output, but it is not yet an agent loop.

### Multimodal

Understands different information sources: text, image, video, audio, or screen state.

### Agentic

Uses perception plus state, goal, action, and feedback to decide what happens next.

**More media is not the same as more agency.**

Durante et al., Agent AI: Surveying the Horizons of Multimodal Interaction, arXiv:2401.03568 (January 2024)

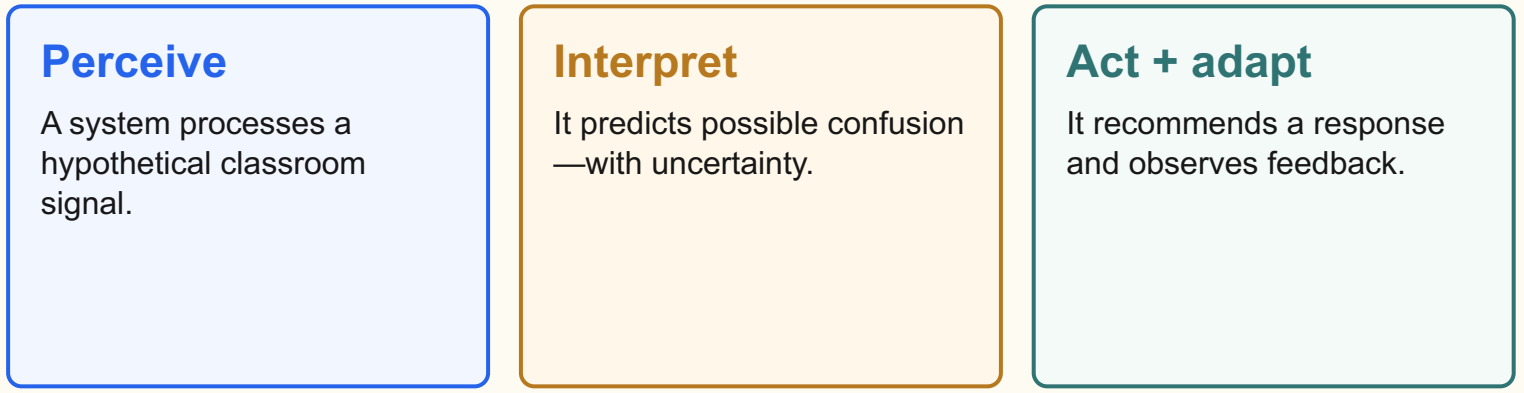
HYPOTHETICAL CLASSROOM TEST

# Description Is Not Yet an Agent Loop

Treat “students look confused” as an uncertain interpretation—not as a reliable fact about students.

## KEY TEACHING POINTS

- Describe a video signal: multimodal perception.
- Choose a reteaching path, quiz, observe, and adapt: agentic loop.
- The interpretation can be wrong before any action occurs.




**Class activity only: do not record or upload real student audio or video.**

REUSABLE AGENT TEST

# The Agent Loop Has Six Questions

Use the same checklist for a robot, game agent, classroom tool, or software assistant.

## KEY TEACHING POINTS

- Trace what enters the system and what it can change.
- Name the feedback signal and the success evidence.
- Evaluate risk before granting authority.

### 1 · Perception
What can it observe?

### 2 · Environment
Where is it grounded?

### 3 · Action
What can it change?

### 4 · Feedback
What happens next?

### 5 · Evaluation
What proves progress?

### 6 · Ethical risk
Who could be harmed?

**Capability and authority are separate design decisions.**

Durante et al., Agent AI: Surveying the Horizons of Multimodal Interaction, arXiv:2401.03568 (January 2024)

ETHICS STAKES

# Wrong Perception Can Become Wrong Action

Signals such as posture, accent, eye movement, lighting, or camera quality are ambiguous—not evidence of intent.

## KEY TEACHING POINTS

- Privacy: screens, faces, voice, grades, and teacher materials.
- Fairness: disability, accent, environment, and unequal technology.
- Accountability: who authorized the action and who corrects it?

### Sensitive signal

Screen, face, voice, grade, or classroom behavior.

### AI interpretation

A probabilistic and potentially biased inference.

### Consequence

Reteach, flag, report, restrict, or change a learning path.

**Higher agency raises the cost of a mistaken inference.**

Durante et al., Agent AI: Surveying the Horizons of Multimodal Interaction, arXiv:2401.03568 (January 2024)

LEARNING BOUNDARY

# AI Helps Learning When Students Still Own the Thinking

The boundary should be observable student understanding—not merely polished final output.

## KEY TEACHING POINTS

- Helpful support can explain, question, plan, revise, and verify.
- Replacement produces the main answer without understanding.
- Students should be able to explain, verify, adapt, and defend.

### Helps learning

The student makes decisions, checks evidence, and can defend the work.

### Replaces learning

The system supplies the reasoning while the student submits the surface output.

**Ask: can the student explain, verify, adapt, and defend the result?**

Durante et al., Agent AI: Surveying the Horizons of Multimodal Interaction, arXiv:2401.03568 (January 2024)

OPERATIONAL POLICY

# “AI Allowed” and “AI Banned” Are Both Too Weak

A usable policy translates ethical language into observable permissions, duties, and boundaries.

## KEY TEACHING POINTS

- Name approved tools and allowed learning uses.
- Name forbidden uses and exam boundaries.
- Require privacy-safe disclosure and human verification.

### Tools

Which systems are approved?

### Uses

What help is permitted?

### Prohibitions

What may not be delegated?

### Disclosure

What assistance must be reported?

**Operational policy replaces slogans with responsibilities.**

Durante et al., Agent AI: Surveying the Horizons of Multimodal Interaction, arXiv:2401.03568 (January 2024)

CLASSROOM GOVERNANCE

# Boundaries Must Come Before Deployment

These are discussion principles for classroom-agent design—not a replacement for institutional policy.

## KEY TEACHING POINTS

- Permission before recording or analyzing people.
- Logs and audits for recommendations and actions.
- Human authority for grades, discipline, accommodations, and appeals.
- No unauthorized upload of course or peer materials.

### Permission

People and data are not automatically available to an agent.

### Audit

Recommendations and actions need a reviewable trail.

### Authority

High-stakes decisions remain with accountable humans.

### Materials

Protect course, peer, and restricted information.

**An agent may suggest; accountable humans must decide and hear appeals.**

Durante et al., Agent AI: Surveying the Horizons of Multimodal Interaction, arXiv:2401.03568 (January 2024)

TAKEAWAY

# From Advice to an Accountable Task Loop

The upgrade is AI participating in perception, action, and feedback—inside explicit human boundaries.

## KEY TEACHING POINTS

- Chatbot: returns advice or content.
- Agent AI: participates in the task loop.
- Governance: who authorized, who checks, and who is accountable?

### Advice

A chatbot proposes what a person might do.

### Task loop

An agent perceives, acts, and receives feedback.

### Accountability gate

A human reviews authority, evidence, risk, and consequences.

**Design the authority boundary—not only the capability.**

Durante et al., Agent AI: Surveying the Horizons of Multimodal Interaction, arXiv:2401.03568 (January 2024)