



Agentic AI **in Higher Education**

Foundations, architectures, pedagogical design and governance of agents in higher education

UNIVERSITY MANUAL · EDITORIAL IAFORTEACHERS

Agentic AI in Higher Education

Fundamentals, architectures, pedagogical design and governance of agents in higher education.



Operational autonomy can be delegated.

Epistemic agency cannot.

Agentic AI in Higher Education. Fundamentals, architectures, design Pedagogical and governance of agents in university education.

Fernando Santamaría González · Editorial IAforTeachers · First edition, 2026

20 chapters · 7 annexes · 73 figures.

ISBN area pending assignment.

Reading index

It can be read from beginning to end, following the arc that goes from concept to practice, or consulted in parts.

I**Conceptual and ontological foundations**

What separates an agent from a conversational system, where the word agent comes from and what learning theories support its use.

14-47**II****Architectures, patterns and coordination**

Anatomy of an educational agentic system and design patterns when several agents work at the same time.

48-73**III****Pedagogical design with agents**

Pedagogical roles, Pedagogically Restricted Autonomy framework, instructional design, and the real limits of customization.

74-105**IV****Applications in higher education**

Tutoring, documented and verified cases, possibilities by fields and agents at the service of academic research.

106-149**V****Reliability, evaluation and governance**

Teaching competencies, how to know if an agent works, systemic security, ethics and institutional governance.

150-204**VI****Perspectives and recommendations**

Towards a reflective agentic pedagogy, future scenarios and emerging lines of research.

205-224**VII****Annexes and apparatus**

The working instruments: glossary, decision matrix, templates, canvas, maturity scale and annotated bibliography.

225-280

Detailed index

Preliminaries	010
CONCEPTUAL AND ONTOLOGICAL FOUNDATIONS	
1 From conversational AI to Agentic AI	015
1.1 What the Agentic AI contributes to university learning: beyond conversational assistance	015
1.2 Differences with traditional educational AI: from conversational assistant to autonomous system	016
1.3 The four distinctive capacities of educational agents	018
1.4 The agentic loop: perceive, plan, act and evaluate	019
1.5 From rule-based systems to agentic architectures: what transition are we witnessing?	021
1.6 Terminological clarification: agent, agenticity, Agentic AI	022
2 The ontological problem of the agent: genealogy and contemporary appropriation	025
2.1 Roots of the concept: from Wiener cybernetics to classical multi-agent systems	025
2.2 Russell and Norvig's rational agent: foundations of classical AI	027
2.3 The Wooldridge and Jennings distinction: weak notion and strong notion of agency	027
2.4 The current semantic reconfiguration: emergent agency versus programmed agency	029
2.5 The risk of agentic illusion: when statistics simulate intention	031
2.6 Operational autonomy and epistemic agency: what can be delegated and what cannot	032
2.7 Implications for education: the difference between instruction and intention	034
3 Theoretical bases of agent-mediated learning	037
3.1 Theoretical frameworks relevant to agentic pedagogy	037
3.2 The agent as mediator, cognitive scaffold and metacognitive mirror	040
3.3 The zone of coagent development: reframing Vygotsky for the algorithmic age	042
3.4 Cognitive delegation and metadelegation: opportunities, risks and the paradox of critical thinking	043
3.5 Cognitive presence: what it means to be present in agent-mediated learning	045

ARCHITECTURES, PATTERNS AND COORDINATION

4	Anatomy and architecture of agentic systems	049
4.1	The cognitive core and the principle of the context window	049
4.2	The system message: where pedagogy enters architecture	051
4.3	Four architectural capabilities and the planner-executor cycle	052
4.4	From words to action: coupling mechanisms	054
4.5	Memory: infrastructure of longitudinal support	056
4.6	Six layers to diagnose an educational agent	057
5	Design patterns and multi-agent coordination	062
5.1	Five canonical patterns as reading glasses	062
5.2	Single agent or multi-agent system	064
5.3	Coordination topologies	065
5.4	Interoperability protocols: from competition to shared infrastructure	067
5.5	The principle of non-universality	069
5.6	Emerging cooperation: promise and warning	070
5.7	The orchestration layer	070
	PEDAGOGICAL DESIGN WITH AGENTS	
6	Pedagogical roles of educational agents	075
6.1	The tutor agent: mediate the process, not deliver the result	076
6.2	The evaluating agent: what it can and cannot do	077
6.3	The instructional design support agent	079
6.4	The simulating agent: controlled professional practice	080
6.5	The epistemic agent provocateur	081
6.6	The coordinator of pedagogical swarms	082
7	The Pedagogically Restricted Autonomy Framework	085

7.1	Why educational autonomy needs its own limits	085
7.2	The five levels	086
7.3	The four questions of orchestrating educator	087
7.4	Application by disciplines and the planned agentic disengagement	089

8 Designing pedagogical instructions 093

8.1	Technical instruction versus pedagogical instruction	093
8.2	Metacognitive instructions	094
8.3	Epistemic Contrast Instructions	095
8.4	Adaptive Scaffolding Instructions	096

9 Personalisation and adaptation: promises and cautions 099

9.1	From differentiation to algorithmic hyperpersonalisation	099
9.2	The four dimensions of agentic personalisation	100
9.3	Three risks, with evidence that has become more precise	102
9.4	The controlled challenge as a design response	103

APPLICATIONS IN HIGHER EDUCATION

10 Agents as intelligent tutors 107

10.1	The four dimensions in tutoring	107
10.2	Adaptive tracking and its limits of interpretation	109
10.3	Formative feedback in three temporalities	110
10.4	When the agent is in the way	111

11 Documented and verified cases 114

11.1	Khanmigo: from mass adoption to the first controlled trial	114
11.2	Pounce: the strongest empirical basis, now with a second replica	117
11.3	PitchQuest: documented architecture, still plausible effectiveness	119
11.4	Cross-sectional analysis: four conditions, and one that has become clearer	120

11.5	Documented failures: survivorship bias and its three cases	122
12	Applications by field	126
12.1	Adaptive tutoring in scientific and technical disciplines	127
12.2	Support for academic writing	128
12.3	Simulation of professional skills	130
12.4	Student guidance and support	132
12.5	Support for teachers in preparation and monitoring	134
13	Agents for academic research	139
13.1	Deep investigation agents: what they are and where they come from	139
13.2	The architectural decision: databases or browser	141
13.3	From static recovery to adaptive strategy	142
13.4	Three tasks, two layers: review, synthesis and research design	143
13.5	Fabricated references: from academic warning to institutional sanction	145
RELIABILITY, EVALUATION AND GOVERNANCE		
14	Teaching skills to work with Agentic AI	151
14.1	Agentic literacy as a foundation	151
14.2	The orchestrating educator: a professional identity in the making	153
14.3	Critical system design: select, configure, monitor	154
14.4	The technopedagogical judgement: the compass of delegation	155
14.5	Professional development levels and training models that work	156
14.6	The new role of the student: the symmetrical half of co-agency	158
15	Evaluation of educational agents: how to know if they work	161
15.1	The bottleneck and trap of easy metrics	161
15.2	From technical test bench to pedagogical evaluation	164
15.3	The evaluating agent and synthetic scenarios: promises with biases that worsen	166
15.4	Reproducibility as a central technical problem	168

15.5	Cost, efficiency, robustness and safety	169
15.6	The operational instrument: five blocks before adopting	170
16	Security and systemic reliability	173
16.1	The distributed anatomy of risk	173
16.2	The catalog of frequent threats	175
16.3	The risks of the action-enabled agent	177
16.4	Multi-agent risks: coordination as an attack surface	178
16.5	The defensive catalogue: honest scope and limits	179
16.6	Lifecycle governance	180
17	Ethics, biases and educational responsibility	184
17.1	Five principles, five design questions	184
17.2	Transparency and explainability: the right to understand	185
17.3	Privacy and data: consent that is not enough	186
17.4	The agentic gap: access, error and benefit	188
17.5	Cognitive dependence as an ethical design flaw	190
17.6	The regulatory landscape: the European regulation and its revised calendar	190
17.7	Responsibility, accountability and the right to human review	191
18	Institutional governance of agentic systems	195
18.1	The implementation roadmap	195
18.2	The three minimum policies	197
18.3	Oversight mechanisms: when governance does not exist	198
18.4	Human supervision in the loop, now dated on the calendar	200
18.5	Teacher training as a prerequisite	202
18.6	Communication with the community	202

PERSPECTIVES AND RECOMMENDATIONS

19	Towards a reflective agentic pedagogy	206
-----------	--	------------

19.1	Current evidence and the hybrid model	206
19.2	The three critical research gaps	208
19.3	The field as early reliability engineering	210
19.4	Seven recommendations for institutions	211
19.5	Seven recommendations for teachers	212
20	Future scenarios and lines of research	215
20.1	The smart campus and its panoptic shadow	215
20.2	Human-AI co-agency in collaborative learning	217
20.3	Metacognitive and self-explanatory systems	218
20.4	The right to agentic silence	219
20.5	Longitudinal evaluation of cognitive impact	220
20.6	The evolution of teaching work: three futures, not one	221
● ANNEXES AND APPARATUS		
A	Glossary of concepts and terms	226
B	Decision matrix for selecting APR level	237
C	Evaluation template for agentic tools for educational contexts	242
D	Questions for institutional reflection before implementing pedagogical agents	250
E	Agent Learning Canvas: visual methodology to design educational experiences with agents	255
F	Application Maturity Scale: Full Framework and Classification Criteria	259
G	Annotated bibliographical references	264
	Index of figures	278

BEFORE STARTING

Preliminaries



Prologue, introduction, note on the method and the author.

Artificial intelligence agents have entered university classrooms before we have a common vocabulary to name them, sufficient evidence to judge them, and agreements to govern them. The promise is compelling: a tireless tutor, immediate feedback, customization at scale. This book turns down the volume on that noise. We've written a guide to exercising sound judgement as technology advances and evidence tries to catch up.

The technical progress of the agents has outpaced the evaluation, **security and governance**. Some applications are already useful. As reliable engineering, however, the field remains young. This tension structures the volume and explains a methodological decision that we have stubbornly maintained: distinguishing what is documented from what is emerging, and what is plausible from what is speculative. When there is evidence, we cite it. When it's missing, we say it.

This manual does not teach programming or celebrate the tool of the week. Nor does it condemn a technology that, used well, can enrich learning. The notions of architecture, coordination and security appear to the extent that they help to decide. The recipient is university teaching staff who already live with these systems, sometimes without having chosen them, and need concepts, *criteria and precautions to act*.

All the chapters return to the same border: the one that separates the **operational autonomy** of the **epistemic agency**. We can delegate the execution of certain tasks to a system, and it is even convenient to do so in some cases. The ability to construct one's own judgment, understand reasons and respond to a conclusion belongs to the person. If we give up that capacity, we empty the act of educating of meaning. Evaluation, feedback, the teaching role, equity and responsibility are variations of this question: what do we give up to the machine and what do we decide to preserve.

We write this book out of conviction and caution. The conviction that agents can amplify learning if they are designed with pedagogical intelligence. The caution that, if used incorrectly, they can atrophy precisely what they claim to enhance. A good part of the future of university teaching is at stake between these two forces, and no tool will decide it for us. We will decide it with the choices we make each day in the classroom. The pages that follow are dedicated to helping you take them better.

Introduction

A chatbot waits for a question and responds. An agent can perceive an environment, plan several steps, resort to external tools, and review what it has done before continuing. This ability to act with a certain autonomy introduces possibilities that conversational AI did not offer and risks that it did not pose either. That's where this manual begins.

Every book on this subject faces the temptation to confuse the possible with the proven. The literature on educational AI abounds with promises that the evidence does not yet support. To keep the categories separate, we apply a **maturity scale** to each relevant statement. What is documented has verifiable public evidence. *The emergent exists, although its scope remains unstabilized.* Plausible is technically possible and has not yet been proven. Speculative projects a reasonable development that may not be fulfilled. The scale tells the reader what ground they are walking on.

From this rigor comes a decision that some might consider excessive. Only three cases deserve, in our opinion, the label of fully documented: Khanmigo, from Khan Academy, Pounce, from Georgia State University, and PitchQuest, from the Wharton School. All three appear analyzed in detail, and with an observation that runs through the book. There is an inverse relationship between the pedagogical ambition of an intervention and the strength of its evidence. The most limited application, Pounce, is the one that has controlled trials. The most ambitious rests on a designer item. *It is worth remembering this every time someone announces a revolution.*

The distinction between operational autonomy and epistemic agency summarizes the position of the manual. The first designates what an agent can do on his or her own: execute, calculate, draft, or organize information. Delegating some of those tasks frees up time and energy. The second belongs to the student: construct knowledge, weigh reasons, doubt and decide. Your delegation disables learning. Debates about evaluation, feedback, personalisation, and governance are organized around this boundary. The pedagogical question is how much should be reserved for the person, even if the machine can do more.

To think about this border precisely, we have developed a set of our own frameworks that the reader will find. The most structural is the Pedagogically Restricted Autonomy framework, which graduates into five levels, from APR-0 to APR-4, how much autonomy should be granted to an agent depending on the task and the student. Notions orbit around it that give a name to what did not have one before. The zone of coagent development reformulates Vygotsky to think about what the student achieves with the support of the agent. The epistemic provocateur describes the one who returns the question instead of the answer. And cognitive presence names the implication of a human thought that accompanies and sustains that of another.

co-agency also has pitfalls. We call **agentic illusion** to the feeling of having learned when a task has barely been completed. Cognitive download can free the mind or atrophy a capacity that stops being exercised. And the agentic gap separates those who use the system to expand their judgment from those who end up replacing it. We propose two specific responses: planned agentic disengagement, which reserves spaces for personal effort, and the coagent transparency framework, which allows us to know who did what. The design decides whether collaboration expands a capability or impoverishes it.

The tour is organized in six parts. The first establishes the conceptual and ontological foundations, what an agent is and what learning theories help us think about it. The second opens the box of architectures, design patterns and coordination between various agents, always at the level that a teacher needs. The third, pedagogical core of the work, presents the roles of the agents, **the APR framework and instructional design**. The fourth descends to the field of applications, with documented cases, plausible possibilities and the emerging line of assisted research. The fifth addresses reliability, evaluation, security and governance, the front where the field is still young. The sixth offers perspectives and recommendations for institutions and teachers.

This manual is designed for those who must decide. It can be read from beginning to end, following the arc that goes from concept to practice, or consulted in parts. A question accompanies both routes: does the use we make of these agents build our ability to think and that of our students, or does it replace it? Each institution and each teacher will have to answer it. The book offers criteria for doing so.

Note on method

A book about agents *artificial intelligence* has an obligation that others do not have, that of saying how it was written. I write this note because the statement that a text has been written by hand is worth little if it is not specified where it was not done that way, and because a reader who is going to rely on two hundred and sixty-five pages on this subject deserves to know at what point the machine intervened.

The original version of the text is mine. So are the thesis that underpins it, the maturity scale, the APR framework, its own concepts and judgments about the quality of the evidence. I am responsible for those decisions and any errors that remain.

In August 2026, I submitted the entire work to an artificial intelligence-assisted editorial review. The intervention focused on style: eliminating repetitions, formulaic constructions, English translations, and overly regular rhythms. The review did not incorporate new cases, data, references or conclusions. When a phrase affected the argument or the degree of certainty of a statement, the final decision was authorial.

This book has been built slowly, step by step, and the tools were used in three moments of that journey, all of them of search and contrast.

To locate literature I used Elicit, Consensus and Perplexity, which allow us to reach works that conventional academic search engines do not return in a field that publishes at this speed. No reference entered this book because it was suggested. All were read before being cited, and those that did not stand up to reading were left out.

To analyze the documents already located I used SciSpace, Scite and Semantic Scholar. The second deserves particular mention, because it shows whether later work supports, *discuss or only mention the one you quote*, and that distinction is what separates a real quote from a decorative quote. To map the field and find what was missing I used Research Rabbit and Connected Papers, which represent the relationships between works and make visible the gaps in a bibliography.

To work on my own sources I used NotebookLM, which only responds to the documents that one gives it and therefore cannot invent what is not there.

To verify the complete bibliographic apparatus I used Claude, with direct queries to the official Crossref and DataCite records. The one hundred and ninety-six entries in this manual were checked one by one, verifying that the identifier resolves and that the title, authors, year, volume and pages match what was declared. The result can be checked reference by reference.

This verification produced eight corrections and it is worth stating it. Seven were precision, and the most significant affected a classic figure that the book gave without the adjustment that subsequent research made to it. Correcting them required rewriting some sentences, so there are a handful of sentences in this manual whose final wording came from that revision process. There are eight passages located and they are documented, not a generated text.

The review did not replace reading the sources or delegate judgment about their quality. No articles were cited from an automatic summary. The classification of each case as documented, emergent, plausible, or speculative remained my decision, made chapter by chapter.

I maintain that this way of working is what this book preaches. It would be incoherent to argue that the ability to verify should never be delegated and to have delegated precisely that when writing it.

About the author

I am a philologist by training, with a degree in Hispanic Philology from the University of León, and I have been working in educational technology for twenty years. I started with the social web when almost no Spanish university was looking at it, between 2006 and 2010.

Much of my work has occurred on the other side of the Atlantic. I have taught at several Colombian universities, I have advised master's projects and I have worked as an advisor on technological matters in Bogotá, and I have toured Argentina, Ecuador, Mexico and other countries giving conferences and accompanying institutions that wanted to change something. I will say it bluntly and without complaint, my work has had more echo in Latin America than in Spain, and I owe to those classrooms a good part of what I know about teaching with technology.

In these twenty years I have seen three or four waves of technological enthusiasm arrive that promised to transform teaching. Almost all left less than they advertised, and the reason was rarely technical. *It was that no one bothered to distinguish what was proven from what only sounded good.*

I wrote this manual because the wave of agents is a different wave, and at the same time it is being counted exactly the same as the previous ones. I wanted a book that made that distinction chapter by chapter, and that a teacher could use to decide without having to believe me. Hence the maturity scale that runs through these pages, and hence I verified the references one by one.

I direct the IAforTeachers Publishing House, where I also publish "The evaluation of learning in Higher Education in times of AI."

This book closes in 2026 and the field does not. Whatever comes next I will publish on my blog, fernandosantamaria.com, where I have been writing for years about learning ecologies and emerging technologies. The articles about agents will continue there, with the same criteria that runs through these pages: **what the evidence ends up confirming**, what denies it and what remains unproven. I appreciate the corrections more than the praise.

PART I

Conceptual and ontological foundations

What separates an agent from a conversational system, where the word agent comes from and what learning theories support its use.

-
- 1 From conversational AI to Agentic AI
 - 2 The ontological problem of the agent: genealogy and contemporary appropriation
 - 3 Theoretical bases of agent-mediated learning

1

From conversational AI to Agentic AI

For almost three years, the conversation about *artificial intelligence* At university it revolved around the same question. What can the model answer? Teachers and students learned to formulate instructions, to refine responses and to live with an interlocutor who, no matter how eloquent he seemed, still depended on someone to initiate each exchange. This comfortable and predictable conversational regime is beginning to prove insufficient.

A different type of system has entered the classrooms without making too much noise. He does not wait for a question to act. Plan multi-step sequences, use external tools, and retain memory of previous interactions. **We call it Agentic AI.** Before discussing its educational uses, we need to specify what it does, how it differs from conversational AI, and what operating cycle sustains its behaviour. Terminology matters, because much of the business confusion begins in names.

A warning about the apparatus in this chapter. Much of what is published today about educational agents are recent proposals with little accumulation of citations. We have preferred to support each statement in the most consolidated work available for each issue, *distinguishing technical evidence about agents*, which already has extensive systematic reviews of the much scarcer pedagogical evidence on its use in higher education. When we only have the second in preliminary form, we say so.

1.1 What the Agentic AI contributes to university learning: beyond conversational assistance

The university does not come to this discussion virgin. Laato and collaborators (2023) were among the first to formulate, for a university audience, what it meant for students to have a competent and available text generator, and they warned even then that the institutional response could not be reduced to prohibition or uncritical adoption. Eager and Brunton (2023) shifted the focus from the tool to the teaching practice, proposing a model of augmented teaching where the system expands the teachers' repertoire without taking their place. The review by Yigci and collaborators (2025) on conversational assistants based on language models in higher education reflects the balance of that first period, with documented benefits in accessibility and availability, and persistent limitations in accuracy, bias and verification.

What comes now is of a different order. A **conversational assistant** He responds and leaves. Let's think about a concrete example. A second-year student asks to explain the difference between internal validity and external validity in an experimental design. The assistant returns a reasonable explanation and the interaction ends there. An educational agent could, on the other hand, maintain an objective throughout the semester, notice that the same concept reappears poorly applied in three successive deliveries and propose, without anyone asking, a sequence of exercises aimed at that specific gap.

Does this provide something that good face-to-face teaching did not already offer? Probably not in essence, but certainly in scale. A teacher with two hundred students enrolled can, in principle, offer longitudinal monitoring that was previously only viable in small groups. *Now, the scale has a cognitive price that should be anticipated.* Essel et al. (2024) observed, in undergraduate students who received immediate responses from a language model, unequal effects on higher-order competencies, with improvements in some tasks and no appreciable improvement in others. The permanent availability of a correct answer does not in itself produce better reasoning.

Regarding the category of educational agent itself, the literature begins to be organized. Kostopoulos and collaborators (2025) position Agentic AI as a class qualitatively different from previous educational artificial intelligence, capable of pursuing complex and long-term objectives with minimal human intervention. Córdova-Esparza (2025), after systematically reviewing eighty-two studies on educational agents published between January 2023 and February 2025, finds a pattern that should be taken seriously: hybrid flows in which teachers curate and moderate the output of the model outperform completely autonomous tutors. Both works are recent and still little cited, so we treat them as reasoned guidance and not as established evidence.

From this follows the epistemic caveat that runs through the entire manual. What technical research on agents documents with increasing solidity does not coincide with what educational research has verified in real classroom conditions. We maintain, and justify it in chapter 11 with a detailed analysis of three cases, that empirical evidence in higher education remains scarce compared to general technical evidence. Khanmigo, from Khan Academy, already illustrates an application with documented evidence today. It constitutes the exception. It is appropriate to establish here, for the first time, the vocabulary that governs the rest of the book. The existence and dissemination of educational agents are a Documented fact. Its pedagogical effectiveness in higher education is, with few exceptions, Emerging. And its equivalence with expert human tutoring is nothing more than Plausible today.

1.2 Differences with traditional educational AI: from conversational assistant to autonomous system

It is important to clarify what we mean by traditional educational AI, because the term conversational assistant has been used for years to designate very different realities. Classic intelligent tutoring systems, such as AutoTutor (Graesser et al., 2005), already incorporated adaptation to the student's pace and dialogue in natural language. They operated, however, within fixed rules, with decision routes predefined by whoever programmed them, and their customization was limited to adjusting the difficulty or sequence within a closed script. The review by Yigci and collaborators (2025) shows that the first generation of assistants based on language models inherited this limitation through another means, that of lacking memory between sessions and of their own initiative.

Where then to draw the border? Kostopoulos and collaborators (2025) propose an operational criterion. A system is considered agentic when it satisfies at least four of six verifiable conditions, including multi-shift planning, persistent memory of progress, autonomous initiative to launch an intervention, and orientation toward a goal sustained over time. The proposal is useful due to its verifiable nature, *although it is not consolidated in the literature*. The most recent technical taxonomies coincide in substance and differ in form, and describe agenticity as a gradation of architectures rather than as a threshold (Xu, 2026, Arunkumar et al., 2026).

The difference exceeds the technical, and this is where the pedagogical nuance decides. A conversational assistant waits. The student initiates, the system responds, the interaction is closed. An educational agent can initiate. Detects a recurring error pattern and takes action before the student asks the question. This reversal of the initiative, from reactive to proactive, is probably the most important change that Agentic AI introduces in university learning. That this change occurs is Documented technically. Its pedagogical scope remains Plausible. *It is also the one that requires greater caution*. An unsolicited intervention raises legitimate questions about student autonomy, and we return to them in Chapter 9.

For those who decide on an institutional purchase, this border has immediate consequences. An assistant without memory or initiative is evaluated with a bank of questions and answers. An agent that plans, remembers and acts requires looking at complete trajectories, not individual responses, and requires asking where the student's history is stored and who can consult it. Chapter 15 develops the evaluation procedures that this difference imposes.

FIGURE
1.1

CHAPTER 1

From conversational AI to Agentic AI

Five behavioral breaks change the relationship between the person and the system

CONVERSATIONAL AI · RESPOND

01 · INITIATIVE

From responding to anticipating

Detects a need and proposes the next step without waiting for an explicit request.

02 · CONTINUITY

From session to process

It preserves useful context and recognizes the trajectory of work over time.

03 · PLANNING

From a response to a sequence

Break down a broad objective into ordered tasks before intervening.

AGENTIC AI · ACT WITH PURPOSE

04 · ACTION

From text to tools

Consult sources, use external systems, and execute authorized operations.

05 · CONTROL

From result to review

Evaluate what has been achieved and correct the course when it does not satisfy the objective.

KEY IDEA The difference is not in conversing better. It is about holding a goal and acting on it.

Figure 1.1. From conversational AI to Agentic AI: five behavioral breaks. Author's own elaboration based on Wang et al. (2024) and Kostopoulos et al. (2025).

1.3 The four distinctive capacities of educational agents

With variations in nomenclature, the technical literature converges on a core of four capabilities that distinguish an agent from a text generator. Wang et al. (2024), in the most cited review to date on autonomous agents based on language models, place them as the constructive core of any agent. Later taxonomies rearrange the pieces without altering the background (Xu, 2026, Arunkumar et al., 2026). Figure 1.2 summarizes these four capabilities applied to the educational context.

The first is the **operational autonomy** and decision making. The agent evaluates options and selects a course of action without human intervention at each step, rather than simply generating the most likely continuation of an input (Wang et al., 2024). In assisted assessment, this means that the system can decide for itself which rubric to apply to an assignment before the teacher reviews the result.

Next comes multi-step planning and reasoning, the skill that has received the most attention. Huang and collaborators (2024) organize the available research into five families, task decomposition, selection between alternative plans, recourse to external modules, reflection on one's own errors and memory support, and show that almost all competent systems combine several of them. Plaat and colleagues (2026), in a systematic review of multi-step reasoning, document both the gains of explicit decomposition and its weakness, the propagation of error between successive steps. Masterman and collaborators (2024) reach a convergent conclusion from the field of architectures. An instructional design support agent who first identifies learning outcomes, then searches for activities consistent with them, and then writes an aligned rubric is executing exactly this capability, with both sides.

The instrumental use of external tools, third of the **four**, converts language into effective action. Chowa et al. (2026) describe this transition as moving from models that only generate text to systems that execute code, query databases, or browse the web to complete a task. In higher education, it enables agents to query the learning management system, verify an actual delivery date, or retrieve updated literature, rather than just what the model memorized during training.

What remains is memory, which is what transforms the pedagogical relationship the most and is also what is least understood. Zhang and collaborators (2025) systematize the matter in the first review dedicated entirely to the memory mechanisms of these systems, and distinguish the internal memory of the context, that which is retrieved from an external store, and that which the agent itself decides to write for future use. The technical origin of the second is well-dated, in the models that consult a massive corpus at the time of generation instead of relying only on its parameters (Borgeaud et al., 2022). *The third is still open ground*, with memory proposals organized by the agent himself that are just beginning to be evaluated (Xu et al., 2025).

An agent who remembers a student's difficulties over the course of a semester can adjust his intervention in a way that no assistant without a persistent memory can match. That same capacity is what turns the student's conversational record into a file that someone guards, with the questions of privacy and data governance that we address in chapter 17. It is a good idea to decide this before hiring, not after.

FIGURE
1.2

CHAPTER 1

The four distinctive capacities of the educational agent

Separate are functions. Combined they form an architecture capable of supporting a task.

01

Operational autonomy

Evaluate options and decide within defined limits, without human intervention at each step.

02

Multi-step planning

Convert a broad goal into concrete and ordered subtasks.

03

Use of tools

Queries systems, retrieves information, and executes authorized actions.

04

Contextual memory

Maintain relevant trajectory and adjust intervention over time.

Agent
educational

KEY IDEA A generator produces content. An agent coordinates capabilities to pursue an educational purpose.

Figure 1.2. The four distinctive capacities of the educational agent. Author's own elaboration based on Wang et al. (2024), Huang et al. (2024), Chowa et al. (2026) and Zhang et al. (2025).

1.4 The agentic loop: perceive, plan, act and evaluate

These four capabilities do not operate separately. *They are articulated in a cycle that the literature describes quite consistently.*, with slight variations in nomenclature, perceive, plan, act and evaluate. Planning is the axis that supports the rest, and it depends on whether the agent decomposes the task into verifiable steps instead of trying to solve it at once (Huang et al., 2024). Masterman and collaborators (2024) observe that designs that clearly define roles and iteratively refine their outputs perform better than those that solve in a single step.

The cycle begins with perception. The agent collects relevant information from the environment, a student's latest submission, a message in the subject forum or a direct query. With that information you enter into planning, where you break down the objective into concrete and ordered steps. The performance executes these steps, often resorting to the external tools from the previous section. The evaluation closes the loop, checks whether the result satisfies the original goal, and when it does not, restarts the process with the additional information generated.

This closure is valid precisely for what it enables, correction. A conversational model generates a response and ends, without checking to see if it solved the asker's problem. An agent evaluating its own output can detect that the first approach was insufficient and try another one without the human telling it to do so. Chowa et al. (2026) document appreciable improvements in anchoring to the environment and in recovery from failures when the orchestration between reasoning and action follows this pattern.

Epistemic honesty requires a warning about the limits of the loop. It works reasonably well on bounded tasks and degrades on long, open-ended tasks, where error in one step contaminates the next, an effect that Plaat and colleagues (2026) identify as the persistent obstacle to multi-step reasoning. Added to this is a verification problem. Xu (2026) points out that the reproducibility of these systems is still limited by the non-determinism of the underlying model and the variability of the tools it invokes, and Arunkumar and collaborators (2026) find that field evaluation practices do not yet allow architectures to be compared reliably.

Transferred to the classroom, the diagnosis becomes operational. An agent can successfully hold a thirty-minute tutorial on a limited concept. On the scale of the book, this solvency in limited tasks is Emergent. Sustaining it for weeks is today, at best, Plausible. It is advisable to closely monitor its performance in support tasks that last for weeks, and above all it is advisable not to delegate to it the final decision on anyone's learning while that reliability cannot be measured.

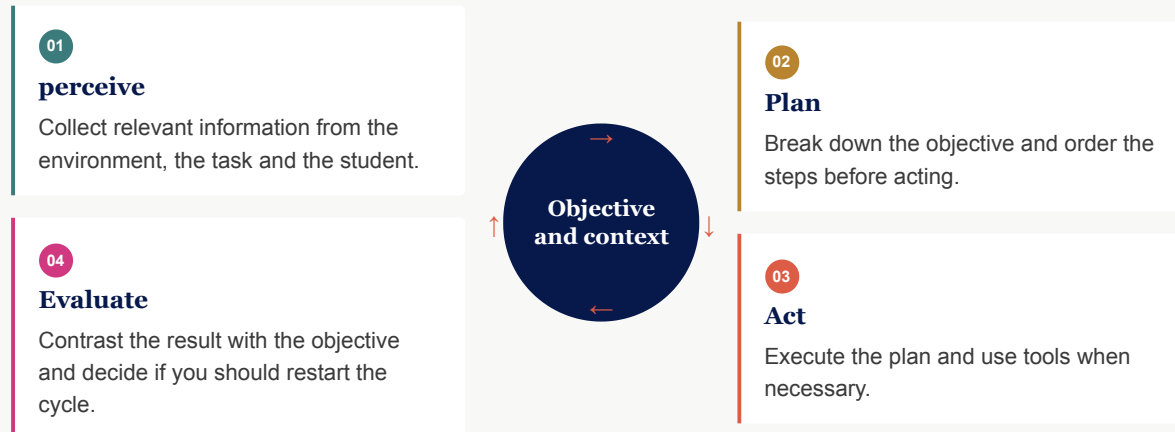
FIGURE

1.3

CHAPTER 1

The agentic loop

Perceive, plan, act and evaluate are not four isolated phases, but a discipline of continuous review



KEY IDEA The decisive feature is the return: if the result is not enough, the system revises the plan.

Figure 1.3. The agentic loop: perceive, plan, act and evaluate. Author's own elaboration based on Huang et al. (2024), Masterman et al. (2024) and Chowa et al. (2026).

1.5 From rule-based systems to agentic architectures: what transition are we witnessing?

The definition of an agent that is recycled today in so many commercial brochures was formulated in 1995. Research in multi-agent systems had already delimited then what it meant to perceive an environment and act on it (Wooldridge and Jennings, 1995). Between that formulation and contemporary agents lie three decades of expert systems with hand-coded decision trees. The idea is old. What is new is the substrate that allows it to be executed at a reasonable cost.

The most recent twist returns, curiously, to the original problem of that tradition. Guo and colleagues (2024) and Li and colleagues (2024) document, in two widely cited reviews, how research has moved from individual agents toward multi-agent coordination. The issues that already concerned agent engineering in the 1990s, division of labor, communication and conflict resolution, then reappear. Chapter 5 develops that turn in detail.

This transition has not been instantaneous or complete, *and it is advisable to resist the temptation to narrate it as a clean break*. For years, rule-based systems coexisted with first approaches of reactive agents, and a good part of the university educational infrastructure continues to operate today with traditional automation, without any agentic component. What has changed in the last two years is the availability of an alternative that did not exist before at an affordable price.

What makes the jump possible? The performance of an agentic system depends on how planning, memory, reflection, and tool use are combined to pursue multi-step goals, rather than on the raw power of the language model underpinning it (Huang et al., 2024, Zhang et al., 2025, Xu, 2026). Accuracy matters to someone who designs an institutional implementation, a topic in Chapter 18. The orchestration and design of the system, and not just the model chosen, determine whether an educational agent achieves the reliability that a university can afford.

FIGURE
1.4

CHAPTER 1

From rule-based systems to agentic architectures

A gradual transition: each stage expands the ability to decide and act

MORE RIGID

MORE AUTONOMY

01

1980S-1990S

Expert systems

Decision trees and hand-coded rules. The behaviour is defined beforehand.

02

YEARS 2000-2010

Smart tutoring

Adapts journeys within predefined decision paths.

03

2022-2023

Generative conversational AI

Responds with fluency and immediate context, but lacks persistent initiative.

04

2024 ONWARDS

Agentic architectures

It integrates reasoning, planning, memory, tools and action within explicit limits.

KEY IDEA More autonomy does not eliminate the previous layers. It integrates them and moves the control center.

Figure 1.4. From rule-based systems to agentic architectures. Author's own elaboration based on Wooldridge and Jennings (1995), Guo et al. (2024) and Li et al. (2024).

1.6 Terminological clarification: agent, agenticity, Agentic AI

We close with a terminological precision that will avoid ambiguities in the rest of the volume. We use three terms that overlap in current use and that should be distinguished with some rigor.

We call an agent, following the classic formulation of Wooldridge and Jennings (1995), a system that perceives its environment, processes that information through some reasoning mechanism and acts on said environment to achieve a goal. This definition, coined almost three decades before big language models existed, remains surprisingly useful, although in Chapter 2 we will show how its application to probabilistic systems introduces complications that its authors did not anticipate.

We use agenticity to mean a gradual property, non-binary, that a system exhibits to a greater or lesser extent depending on how it combines autonomy, proactivity, planning capacity and persistent memory. A system is located at a point on a spectrum. The available technical taxonomies order this spectrum by architecture rather than by threshold, which supports this gradual reading (Xu, 2026, Arunkumar et al., 2026). Graduality avoids the mistake, common in technological dissemination, of labeling as an agent any product that incorporates a conversational assistant with some added function.

We reserve the term Agentic AI for a contemporary architecture: a large language model embedded in a software loop that schedules, invokes tools, and preserves memory between interactions. The agentic capacity arises from that construction, not from the isolated model. The nuance sustains the transversal distinction of the manual. We can delegate operational autonomy within certain limits. *Epistemic agency requires a different degree of caution*, because it compromises the formation of the student's own judgment. The next chapter examines that difference in more depth.

● THREE TERMS, THREE LEVELS

Agent: system that perceives, reasons and acts on its environment to achieve a goal (Wooldridge and Jennings, 1995). *Agenticity*: gradual property that combines autonomy, proactivity, planning and memory. Agentic AI: Architecture that integrates a large language model into a software loop capable of planning, tooling, and remembering.

CHAPTER SUMMARY

Four capabilities separate the agent from conversational assistance: operational autonomy, multi-step planning, *use of external tools and persistent memory*. Together they form a loop of perception, planning, action and evaluation that allows course correction without asking for human approval at every step.

That loop loses reliability in long, open tasks. We also do not yet have evaluation practices that allow us to compare architectures with sufficient solvency. The evidentiary asymmetry is clear: technical evidence is growing rapidly, *while pedagogical evidence in higher education remains scarce*. The available data favors hybrid flows with teaching supervision.

The terminology is fixed. An agent perceives, reasons and acts. Agenticity is a gradual property. The *Agentic AI* It describes an architecture, not an isolated model quality. With this basis, the next chapter examines in what sense a statistical system can be called an agent.

References

20
TICKETS

Arunkumar, V., G. R., G., & Buyya, R. (2026). Agentic artificial intelligence (AI): Architectures, taxonomies, and evaluation of large language model agents. arXiv. <https://doi.org/10.48550/arXiv.2601.12560>

Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., van den Driessche, G., Lespiau, J.-B., Damoc, B., Clark, A., de las Casas, D., Guy, A., Menick, J., Ring, R., Hennigan, T., Huang, S., Maggiore, L., Jones, C., Cassirer, A., & Sifre, L. (2022). Improving language models by retrieving from trillions of tokens. Proceedings of the 39th International Conference on Machine Learning, 162, 2206–2240. <https://proceedings.mlr.press/v162/borgeaud22a.html>.

- Chowa, S. S., Alvi, R., Rahman, S. S., Rahman, M. A., Raiaan, M. A. K., Islam, M. R., Hussain, M., & Azam, S.** (2026). From language to action: A review of large language models as autonomous agents and tool users. *Artificial Intelligence Review*, 59. <https://doi.org/10.1007/s10462-025-11471-9>
- Córdova-Esparza, D.-M.** (2025). AI-powered educational agents: Opportunities, innovations, and ethical challenges. *Information*, 16(6), 469. <https://doi.org/10.3390/info16060469>
- Eager, B., & Brunton, R.** (2023). Prompting higher education towards AI-augmented teaching and learning practice. *Journal of University Teaching and Learning Practice*, 20(5). <https://doi.org/10.53761/1.20.5.02>
- Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O.** (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence*, 6, 100198. <https://doi.org/10.1016/j.caeai.2023.100198>.
- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A.** (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612-618. <https://doi.org/10.1109/TE.2005.856149>
- Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., & Zhang, X.** (2024). Large language model based multi-agents: A survey of progress and challenges. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 8048-8057. <https://doi.org/10.24963/ijcai.2024/890>
- Huang, X., Liu, W., Chen, X., Wang, X., Wang, H., Lian, D., Wang, Y., Tang, R., & Chen, E.** (2024). Understanding the planning of LLM agents: A survey. *arXiv*. <https://doi.org/10.48550/arXiv.2402.02716>
- Kostopoulos, G., Gkamas, V., Rigou, M., & Kotsiantis, S.** (2025). Agentic AI in education: State of the art and future directions. *IEEE Access*, 13, 177467-177491. <https://doi.org/10.1109/ACCESS.2025.3620473>
- Laato, S., Morschheuser, B., Hamari, J., & Björne, J.** (2023). AI-assisted learning with ChatGPT and large language models: Implications for higher education. *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)*, 226-230. <https://doi.org/10.1109/ICALT58122.2023.00072>
- Li, X., Wang, S., Zeng, S., Wu, Y., & Yang, Y.** (2024). A survey on LLM-based multi-agent systems: Workflow, infrastructure, and challenges. *Vicinagearth*, 1. <https://doi.org/10.1007/s44336-024-00009-2>
- Masterman, T., Besen, S., Sawtell, M., & Chao, A.** (2024). The landscape of emerging AI agent architectures for reasoning, planning, and tool calling: A survey. *arXiv*. <https://doi.org/10.48550/arXiv.2404.11584>
- Plaat, A., Wong, A., Verberne, S., Broekens, J., van Stein, N., & Bäck, T.** (2026). Multi-step reasoning with large language models, a survey. *ACM Computing Surveys*, 58, 1-35. <https://doi.org/10.1145/3774896>
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.** (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186345. <https://doi.org/10.1007/s11704-024-40231-1>
- Wooldridge, M., & Jennings, N. R.** (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2), 115-152. <https://doi.org/10.1017/S0269888900008122>
- Xu, B.** (2026). AI agent systems: Architectures, applications, and evaluation. *arXiv*. <https://doi.org/10.48550/arXiv.2601.01743>
- Xu, W., Liang, Z., Mei, K., Gao, H., Tan, J., & Zhang, Y.** (2025). A-MEM: Agentic memory for LLM agents. *arXiv*. <https://doi.org/10.48550/arXiv.2502.12110>
- Yigci, D., Eryilmaz, M., Yetisen, A. K., Tasoglu, S., & Ozcan, A.** (2025). Large language model-based chatbots in higher education. *Advanced Intelligent Systems*, 7, 2400429. <https://doi.org/10.1002/aisy.202400429>
- Zhang, Z., Dai, Q., Bo, X., Ma, C., Li, R., Chen, X., Zhu, J., & Dong, Z.** (2025). A survey on the memory mechanism of large language model-based agents. *ACM Transactions on Information Systems*, 43, 1-47. <https://doi.org/10.1145/3748302>

END OF THE FREE SAMPLE

You have read the complete opening chapter of Agentic AI in Higher Education. The full handbook continues from concepts and architectures to pedagogical design, evaluation, safety, ethics and institutional governance.

THE COMPLETE EDITION INCLUDES

- 20 chapters and 7 practical annexes
- 73 original figures and 17 tables
- A 322-source bibliography
- The 24-page Agentic AI Teaching Toolkit
- PDF and offline HTML editions

CONTINUE WITH THE COMPLETE BOOK

<https://payhip.com/b/UOPxr>