Where does AI training really lead?

What changes when a learned capability enters work, study and private life.



Gruppo Radar · Comprendere oggi. Costruire domani
by effettoantropico.life

A person types a question into the dialogue window of what looks like a chatbot (a system with which one converses in natural language). A few seconds later, they receive an orderly, plausible answer, perhaps even better than they expected. They copy it into a document and use it to prepare a meeting, study a subject, clarify a thought or organize something that seemed confused only a moment before.

The action is simple. The result is not.

That answer may be correct but not necessarily suited to the circumstances. It may be an excellent first draft, but if it is not validated, it can become a very poor final decision. It can certainly save time today but, if not used with discernment, weaken the ability needed tomorrow. It can help someone who is learning, create doubts for someone who already knows their work, or make ideas more uniform when, without the same initial suggestion, they would have taken different paths. With this issue I close the trilogy devoted to one of the most important processes in artificial intelligence: training. Anyone who uses these tools should understand what it means, because capabilities, limits and possible applications depend on it.

What does it really mean to train an AI? What is it for? And why is it so important?

The first issue showed who and what continue to teach the machine. The second followed a capability through a controlled test, to understand what becomes observable when the system is placed under pressure. This third issue takes the step closest to our experience: it follows that capability as it reaches a desk, a classroom, a laboratory, a sensitive profession or life at home.

We will not find a single answer. We will find something more useful: the conditions that change the direction of the result. The task. The person's experience. The way the tool is introduced into the work. The ability to check what it produces. The measure by which we decide whether something has really improved.

> 🧠 Radar Insight Training builds a capability. Value arises afterwards, when someone assigns it a task, checks the proposal and decides what to do with it. Between the model and the result there is always a human system.

1. The correction that looks like teaching

When we correct an answer and the next one improves, it is natural to think that we have taught the model something. In many cases, however, what we see is closer to good conversational continuity than to new training.

The words exchanged remain for a time in the context window, the space of text the system can consider while preparing its next answer. The prompt, the instructions we

write, can therefore guide its immediate behavior without changing the parameters, or weights: the numerical values adjusted during training and distributed throughout the model's network [E01].

This does not make the correction useless. It may be exactly what is needed to obtain a more precise text, enforce a constraint or bring the reasoning back onto the right path. Its effect may simply remain within that conversation.

In other cases, a product memory comes into play, retaining certain information and bringing it back later. Or the system consults an external archive through RAG, the English abbreviation for generation that first retrieves relevant documents and then uses them as context [E03–E04]. In a company, a correction can be collected in a dataset, a structured collection of examples, and become material for fine-tuning, further targeted training. It can also be used to evaluate the system or prepare a future version [E02, E05].

These are different outcomes. A memory is not a new model. A retrieved document is not automatically incorporated into the weights. A user's comment does not become training simply because it was written. For the system to change durably, an identifiable choice, process and responsibility are needed.

This technical distinction has a very concrete consequence. If information is needed only for the current task, keeping it for a long time may be useless or even inappropriate. If, instead, it is to become part of an organization's shared knowledge, leaving it buried in one person's chat is fragile. And if it is to change a tool's behavior durably, hoping that it "has understood" is not a method.

The first skill, then, is not finding a brilliant formula. It is knowing what we want to build: a note, a draft, a verified decision, a searchable archive, a repeatable procedure, an ability that must remain in the person's hands and not be delegated to the tool, or one of the many other possible outcomes, including those that can strengthen individual autonomy. Before asking artificial intelligence how to achieve it, we must know what should remain after the conversation and to whom it should belong.

2. From intention to what remains

In the everyday use of artificial intelligence, we see a flow with a beginning and an end, but it is in the middle that almost everything is decided.

The path begins with the person's intention. Not with the machine. Someone wants to understand, create, choose, remember, organize or act. They then provide context: information, sources, examples, limits. The system interprets that material through

what it has learned and puts forward a proposal. At that point the task returns to the human side: the answer must be checked, modified, accepted or rejected. Only then can it become a real action.

We can represent the cycle in nine steps:

1. the person's intention;
2. context provided;
3. the system's interpretation;
4. proposal or generation;
5. verification;
6. human decision;
7. implementation or action;
8. feedback and correction;
9. subsequent learning.

The list is a map, not a mandatory scheme. In real life some steps intertwine and others are skipped. That is precisely the point. A copied answer bypasses verification. An automatic process without records makes it difficult to understand what happened. A correction left in the conversation may teach the person a great deal and the system nothing.

The final step also branches. The person can learn, recognizing an error or building a better method. The organization can learn, updating a procedure, an archive or the assignment of responsibilities. The system, by contrast, changes durably only if that feedback actually enters a memory, a configuration or a new governed training cycle.

The same capability can also take on different roles. Sometimes it is an occasional answer. Sometimes it is a copilot, because it accompanies a person who continues to steer. It can become automation, the repeated execution of defined steps, or an agent, a system capable of choosing and chaining several actions using the tools and information available. The degree of operational autonomy, however, does not by itself measure value. A single answer can address a specific need well; an agent can multiply an error before anyone manages to see it.

To understand whether its use worked, a counterfactual is needed: a credible comparison with what would have happened without that tool. We must then choose

what to measure: time, quality, quantity, learning, well-being. And we must ask whether the benefit holds when the task changes or assistance is removed.

Without these comparisons, “useful” becomes a word that is too broad. An email written sooner, a better decision and a less fragmented day can coexist, but they are not the same result.

3. Work has no universal percentage

Studies of work tell a story less linear than commercial promises and richer than dismissals made in advance. In some tasks artificial intelligence produces measurable, observable gains. In others it adds checking, distraction or rework. This is not a contradiction: it is the effect of different kinds of work entrusted to different people.

A randomized study, in which access to the tool is assigned randomly to make the comparison more credible, involved 444 professionals in short writing tasks. Average time fell from about 27 to 17 minutes and assessed quality rose [E06]. At a large customer support company, 5,172 workers resolved on average about 15% more problems per hour. The gain was concentrated mainly among less experienced workers [E08].

Positive results have also been observed in software development. Three company experiments, involving a total of 4,867 developers, recorded more completed tasks among those with access to a coding assistant [E09]. But the number does not tell us what happens afterwards: maintenance, security, hidden errors and technical debt (the future cost of postponed corrections and maintenance) were not captured by the same measure.

The most instructive point emerges when the task meets the model’s frontier, the irregular boundary between what a particular configuration can handle well and what challenges it. It is called the *“jagged frontier”* precisely because two apparently similar activities can fall on opposite sides. In the experiment with 758 consultants at BCG, participants worked faster and achieved better results on tasks within that frontier. On a task outside it, however, the probability of giving the correct answer fell by 19 percentage points [E07].

Human competence therefore does not consist only in knowing how to obtain an answer. It also consists in recognizing when confidence in the form conceals a weakness in the content.

A study by the research organization **METR** makes another possible illusion visible. Sixteen experienced developers worked on 246 real problems in code repositories, the

structured spaces where programs and changes are stored. With the tools available in early 2025, they took about 19% more time. Yet before the study they expected substantial time savings and, even afterwards, continued to believe they had been faster [E10].

Even when a local gain exists, the organization may change much less. In 66 companies, a tool integrated into email, meetings and writing reduced by about two hours per week the time spent on email and after-hours work. No changes emerged, however, in the overall amount or composition of activities [E11]. In Denmark, the first years of adoption brought new tasks and reorganizations, without large detectable average effects on earnings and hours worked [E12].

Time made available has no automatic destination. It can become greater care, more analysis, less evening work. It can also be immediately filled with new demands. That is why the interesting question is not how much AI *"yields"* in general, but which work it removes, which it shifts, which checks it introduces and who truly receives the benefit. Above all, one question remains: *how much of that operational and professional advantage can a person really turn into time of their own, autonomy and quality of life? And how much is instead absorbed back into work because we no longer know how to be without it?* The answer depends on more than the tool or the organization. When interests, relationships, curiosity, rest and personal plans are no longer cultivated, time made available risks being perceived as a void and quickly filled with more work. Artificial intelligence can give us time back, but it cannot decide for us what we will do with it or automatically teach us how to live it.

📌 **Did you know?** In the METR case, perception and measurement pointed in opposite directions: the developers felt faster, while recorded time indicated a slowdown. A smooth experience is not proof of productivity.

4. Studying well does not simply mean finishing sooner

In learning, the distance between performance and ability becomes impossible to ignore. An exercise can be correct even when the person submitting it could no longer complete it without assistance.

Nearly a thousand secondary-school students in Turkey studied mathematics with either freer access to a model or a tutor equipped with pedagogical constraints. During practice, results improved in both groups, and even more with the tutor. In the subsequent test, taken without AI, those who had used the basic version scored about

17% below the control group. The constrained tutor reduced that harm, without producing a clear advantage in the unaided test [E13].

The result does not say that AI should be kept out of the classroom. It says that it must be given an educational role. Tutor **CoPilot**, tested with 900 tutors and 1,800 students from disadvantaged communities, suggested strategies to the adult while working with the student. The rate of passing short assessments rose from 62% to 66%, with a greater benefit for tutors who were initially less effective [E14]. The machine did not hand the solution to the student: it helped the tutor formulate more useful questions and avoid replacing the student's own effort.

In a university physics course, a tutor built around pedagogical principles produced, over two lessons, better results than a specific form of active learning, teaching based on student participation rather than explanation from the front alone. The median time spent was also lower [E15]. But the treatment was not a generic chatbot. It was a designed environment, with selected content, an individual pace and a structure consistent with the educational objective.

These experiments do not describe a single educational effect of AI. Rather, they show that the tool inherits the method we build around it. It can offer a shortcut that avoids necessary effort or support that makes that effort more productive.

The same applies to adults who study. Asking for an explanation, being tested, comparing two interpretations and building a revision plan are different uses from having the result handed to us. More mature use does not eliminate all effort: it protects the effort that produces understanding and reduces the effort that adds no value.

⚠️ **To watch** If we measure only the exercise completed with AI, we risk calling a successful delegation learning. The decisive question is what the person can do afterwards, when the tool is absent or the problem changes.

5. When generating becomes easy, choosing matters more

For writing, designing and research, artificial intelligence lowers the cost of a first proposal. In a few seconds it can offer titles, images, variants, code and hypotheses. At that point scarcity shifts: initial material is no longer missing. What is scarce is the attention to compare it, the judgement to discard it and the time to verify what deserves to remain.

An experiment on writing short stories showed this double movement. Access to generated ideas improved average individual ratings of creativity, quality and enjoyment, especially among people starting at a lower level. The stories, however, became more similar to one another [E16]. Individual improvement and diminished collective variety can both be true at the same time.

There is no need to attribute intention to the machine. Convergence arises from statistically plausible and familiar proposals. If the system always enters before the person has sought a direction of their own, the first idea received can become an anchor. If it comes afterwards, to test a choice or seek alternatives, it can broaden the comparison.

Scientific research offers a different case. FunSearch does not mistake the model's first answer for a discovery. It generates many small programs, runs them, eliminates those that do not work, keeps the best and starts again. It has produced new mathematical constructions for problems where efficient verification and a score capable of guiding selection existed [E17]. The result comes from the entire cycle, not from a particularly convincing sentence.

Not every field has an automatic verifier. A biological hypothesis requires experiments. A social claim requires definitions and data. A historical interpretation requires sources, context and comparison. Where checking is costly, ease of generation can increase noise faster than our ability to eliminate it.

A large observational study, which analyses what happened without randomly assigning AI use, examined 41.3 million scientific papers. The authors found an association between AI tools and more publications and individual citations; at the same time, they measured a reduction in the collective breadth of topics and interaction among researchers [E18]. Precisely because it is observational, the study does not establish causality: it does not prove that AI produced those careers or results. It does, however, make visible a tension between individual incentive and diversity in research.

The quality of sources is part of the same story. A preprint, a scientific manuscript circulated before peer review, had attracted attention with striking figures about researchers' productivity. It was withdrawn. MIT stated that it lacked confidence in the provenance, reliability and validity of the data and recommended against using it [X01]. A number does not become true because it confirms a hope or a fear.

🧠 **Radar Insight** AI makes it less costly to produce candidates. That is precisely why the choice of problem, the quality of the test and the courage to discard almost everything gain value.

6. When the answer touches a sensitive profession

Medicine allows us to see with particular clarity three levels often merged in debate: what the model can do on its own, what a person can do with its help and what actually happens to the patient.

A pre-registered study, in which the main questions and analyses are stated before the results are observed, involved 1,298 British adults and ten medical scenarios. Three models, when directly given the complete case, identified at least one relevant condition in 94.9% of trials. In 56.3% they also indicated the correct level of care to seek, from self-care to calling an ambulance. People using the same models, by contrast, remained below 34.5% in identifying conditions and below 44.2% in choosing the level of care, without doing better than the control group [E19].

Another experiment, conducted between November 2023 and April 2024, involved 92 doctors and five simulated clinical cases derived from real, anonymized encounters. The group assisted by GPT-4 — a model that now belongs to an earlier generation — achieved a mean clinical management reasoning score 6.5 percentage points higher, while taking about two minutes more per case [E20]. This is a benefit observed in the task studied and relating to that specific generation of model: it cannot automatically be transferred to current systems. Nor does it demonstrate that patients received better care, because no real patients were involved and clinical outcomes were not measured.

Earlier research with radiologists and internal and emergency medicine physicians shows how fragile formal oversight can be. When the suggestion was inaccurate, accuracy worsened whether it was presented as produced by AI or attributed to a human being [E21]. The mere presence of a professional does not guarantee effective supervision. Time, information, expertise and a real opportunity to disagree are needed.

The same principle applies in law, finance, public administration and any activity capable of changing rights or material conditions. The system can search, synthesise and propose. Responsibility remains with those who decide which tool to use, which data to provide, which risk to accept and what remedy to offer when something goes wrong.

⚠️ **To watch** A model that answers a complete case well may not help someone who must decide what information to provide, interpret the advice and decide whether to follow it. Interaction is not a detail: it is part of safety.

7. In private life we know what goes in, not yet what remains

Everyday use does not stop at the office. The British statistics office found that, during the previous twelve months, people had knowingly used AI for work or education, household activities and hobbies or interests. A third, by contrast, said they had not knowingly used it [E22].

Telemetry, the automatic recording of uses observed within a product, offers another perspective. In **ChatGPT** data, about seven in ten messages in services aimed at individuals were classified as non-work-related; practical guidance and writing appeared among the main uses [E23]. The source is the provider itself: it describes its own product and does not demonstrate a benefit for the population.

Behind those categories are very different activities. A person may use conversation to put a thought in order, prepare a difficult communication or turn a confused problem into more precise questions. They can build an outline, organize notes, compare options or spread a project over time. They can ask for help understanding the structure of a document and arrive better prepared to see a doctor, lawyer, teacher or adviser. This does not replace the professional: it can improve the question put to them.

The tool can accompany language practice, suggest variations in tone, describe an image, transcribe or summarize audio. It can generate texts, images and sounds to select and refine. It can adapt content to a different reading level, turn writing into speech or help construct an alternative description. For someone facing a visual, linguistic or organizational barrier, moving between formats can be a concrete and important capability.

But an observable capability is not yet a demonstrated benefit. A summary may make a document accessible or erase the very exception that matters. A plan may free attention or produce an unrealistic schedule. A conversation may clarify thinking or politely confirm a mistaken premise. A translation may open a relationship or flatten a cultural meaning. A generated image may expand expression or replace the search for a voice of one’s own.

A US survey shows that many people would accept some form of everyday assistance, while a majority feel they have little or no control and would like more [E24]. It is not a measure of dependence. It does show, however, that adoption and autonomy do not necessarily advance at the same pace.

The technical location of the interaction also changes what is delegated. In data from the provider Anthropic, direct use by people appears more collaborative; in the API, the interface through which other software calls the model without going through the ordinary chat, automation dominates [E25]. A suggestion visible to the user is different from a decision taken behind an application. And many people already encounter embedded systems (recommendations, assistants, recognition systems) without always identifying them as AI [E26].

Here the limits of the evidence must remain in the foreground. We know that AI is used in private life and has useful capabilities in concrete activities. We do not yet have sufficient data to claim that it durably improves personal organization, relationships, autonomy or well-being. Robust series on dependence and on what happens to skills after years of delegation are also lacking.

Prudence does not forbid these uses. It only prevents us from turning them into a promise. When we ask for help with a personal choice, we must know what we are not delegating: values, acceptable risk, responsibility towards others and the right to change our minds.

8. Governing the tool is a form of literacy

For a long time digital competence was associated with the ability to use a device or a program. With generative artificial intelligence, the threshold shifts. The interface appears simple precisely because much of the complexity remains hidden.

The new literacy is not a collection of formulas for writing perfect prompts. It begins earlier, when we define the result we need. It continues with choosing the context to provide, avoiding the disclosure of sensitive data or unnecessary information. It requires assigning the tool a role: generator, critic, tutor, researcher in an archive or executor of controlled steps.

Then comes verification. An original source, a repeatable calculation, a test without assistance, the judgement of a knowledgeable person. And then comes the full cost: not only the minutes saved, but those spent checking, correcting and recovering from an error.

Finally, we must recognize the frontier. A system can be excellent at rephrasing and fragile at inference; useful in proposing alternatives and unreliable at confirming a rare fact. That frontier changes with the model, the tools and the context. It cannot be learned once and for all.

Control becomes real only when we can refuse. If ignoring the suggestion is too costly, if no one can reconstruct the information that produced it or if the organization punishes dissent, human presence becomes a signature added at the end.

Closing the cycle also means knowing who has learned. The person may have acquired a method. The organization may have updated a process. The system may not have changed at all. Keeping these three forms of learning separate prevents us from attributing to the machine what arose from human experience and from losing what that experience might teach the organization.

This is not a statistical scale of skills. It is a grammar drawn from the cases observed. Those who use AI well are not those who accept more answers, but those who know how to turn some into results, discard many and explain why.

9. When error acquires mass

So far we have followed capabilities that produce words, code, images, suggestions and decisions. There is a further step: connecting the system to a body, a machine or an infrastructure. Here the generated content does not remain information alone. It can become movement.

The US standards institute, NIST, describes physical AI as a system in which the software that interprets data and produces commands, the robot that translates them into action, the task and the environment must be evaluated together. It also points to a significant distance between academic research and what can practicably be implemented in industry [E27]. A successful demonstration does not automatically encompass durability, maintenance, environmental conditions, failures and the quality of human intervention.

The discipline required is not new. It is the same as we encountered in writing, study and medicine, with a higher cost of error: define the task, limit permissions, record what happens, measure the result and make it possible to stop.

The European AI Regulation, for the high-risk systems to which it applies, describes human oversight that does not amount to the mere presence of a person. Those in control must be able to understand capabilities and limits, interpret the proposal, disregard or override it and safely stop the system. They must also recognize automation bias, the tendency to trust an automatic suggestion because it appears authoritative [E28]. Guidance from the US authority for medicines and medical devices also emphasizes, within its own scope, the professional's ability to review the

basis of a recommendation [E29]. It is a condition of accountability, not proof of effectiveness.

For the everyday spread of physical AI, the data remain insufficient. We do not have comparable global denominators, consistent error rates or independent measures of human intervention. For this reason the physical world remains the horizon of the dossier, not its centre. It makes a consequence visible: when a learned capability is connected to an actuator (the component that turns a command into movement or physical action) the quality of governance becomes material.

10. Direction is not written in the weights

Training does not lead to a single place or draw an inevitable future. It builds possibilities that are then selected, integrated and directed by people and organizations.

It can lead to a draft that frees time to think better. To a tutor that helps the adult avoid giving the answer in the student's place. To a research environment capable of generating many hypotheses and discarding them using a verifiable criterion. To a clinical recommendation that improves an exercise but does not yet demonstrate better care. To an everyday assistant that organizes material or makes content more accessible, without our yet being able to measure its long-term effect on autonomy.

It can also lead to unjustified confidence, skills left unpracticed, code that takes more time than it seems, ideas that are individually better but collectively more uniform. None of these outcomes is contained in the word "intelligence". It arises from the meeting of capability, task, person, interface, incentives and control.

The trilogy's concluding question is therefore not *"what will AI do?"*. It is too large and, precisely for that reason, of little use.

We can ask more concrete questions. *What capability are we introducing? At what point in life or work? What must remain with the person? How will we recognize a good result? Who can correct, challenge or stop it? Who is accountable when a proposal becomes action?*

The decisive boundary is not between those who use artificial intelligence and those who do not. It lies between a use that accumulates answers and one that builds results that are legible, controllable and attributable to people. The machine brings capability. Direction, judgement and responsibility remain ours.

Methodological note and traceability

Provider sources describe only the products observed and are not independent. Surveys describe use and perceptions, not benefits. Experiments support causal conclusions only for the task, population and duration studied. Data remain insufficient on causal benefit in private life, dependence and long-term skills, overall organizational transformation, general scientific productivity, long-term clinical outcomes, full costs, unrepresented geographies and safe deployment of physical AI.

Essential sources

- Brown et al., *Language Models are Few-Shot Learners* — Primary source
- Noy e Zhang, *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence* — Primary source
- Dell'Acqua et al., *The Jagged Technological Frontier* — Primary source
- Brynjolfsson, Li e Raymond, *Generative AI at Work* — Primary source
- Cui et al., *The Effects of Generative AI on High-Skilled Work* — Primary source
- METR, *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity* — Primary source
- Bastani et al., *Generative AI can harm learning* — Primary source
- Wang et al., *Tutor CoPilot* — Primary source
- Kestin et al., *AI tutoring outperforms in-class active learning* — Primary source
- Doshi e Hauser, *Generative AI enhances individual creativity but reduces the collective diversity of novel content* — Primary source
- Romera-Paredes et al., *Mathematical discoveries from program search with large language models* — Primary source
- Bean et al., *Reliability of LLMs as medical assistants for the general public* — Primary source
- Goh et al., *GPT-4 assistance for improvement of physician performance on patient care tasks* — Primary source
- ONS, *Public opinions and social trends, Great Britain: June 2026* — Primary source
- OpenAI, *How People Are Using ChatGPT* — Primary source

- Regolamento (UE) 2024/1689 — Primary source

© 2026 Barbara Gattobigio · Gruppo Radar

Comprendere oggi. Costruire domani · by effettoantropico.life

All rights reserved on original content. Free access does not authorize reproduction or republication without consent, except for uses permitted by law. Third-party materials remain the property of their respective owners. To share this issue, use the link to the article.