

Redefining Human-in-the-Loop in the Agentic AI Landscape: Reading the Map of the Moment in a Shifting Labyrinth

Sefa Emre Öncü, Aras Bozkurt, Ramesh Chander Sharma

Abstract: Human-in-the-loop is treated as a safeguard in automated systems. Nevertheless, in agentic systems the ground for that safeguard is becoming contestable. With agents that split the goal, call tools and screen their own output, the loop ceases to be singular and a shifting labyrinth appears, one whose walls do not stay in place. How long a person remains in the loop can be measured, yet that measure does not show whether oversight has taken place. What is decisive is the aptness and the efficacy of the intervention. Aptness concerns both the layer an intervention lands on and the moment it arrives. Efficacy in turn addresses whether the outcome would have differed without it, and whether it differed for the better. The literature discussed here is read under three themes: where and when oversight happens, the illusion of oversight, and the competence required. As the interface grows simpler it hides more of what runs beneath it, and unless transparency grows with it the expertise oversight demands grows too. The hardest part left to the person then settles behind an appearance of ease. Oversight competence is gained by being able to read the design; its cost falls at the front and its return appears on every run. Oversight should be documented not by time spent in the loop but by a record of the aptness and the efficacy of interventions. Distributing oversight across defined oversight functions rather than a single person in multi-layered arrangements is offered as a research agenda.

Keywords: agentic AI, human-in-the-loop, human oversight, aptness of intervention, efficacy of intervention, human-agent interaction, levels of autonomy, design literacy

Highlights

What is already known about this topic:

- Human-in-the-loop is a half-century-old concept, grown out of work on levels of automation and supervisory control, and it now covers uses from data labelling to decision support. The sense at issue here is oversight of deployed agentic systems.
- Automation takes over what can be automated; what is left to the person is the work the designer could not think how to automate, and the very work they now do least often.

What this paper contributes:

- A displacement of the criterion: the indicator of oversight is not presence alone but the aptness and the efficacy of the intervention.
- A relation that design can change: as the interface grows simpler and hides more of what runs beneath it, oversight demands more expertise unless transparency grows too.

Implications for theory, practice and/or policy:

- Reporting oversight through a record of the aptness and efficacy of interventions rather than through time spent present.
- Widening AI literacy so that it covers being able to read agentic systems.
- Distributing oversight in multi-layered agentic systems across defined oversight functions rather than a single person.

Introduction

You build the world of the dream. We take the subject into that dream, and let him fill it with his subconscious.

Cobb, to the architect. Nolan, Inception (2010).

In Nolan's (2010) Inception, the problem is not the depth of the dream but being caught inside a world one has built oneself. The movie calls this limbo: unconstructed dream space, holding nothing but what was left behind by those trapped there before. The one who can leave is the one who knows how the structure was put together.

Something similar can be said of where the human stands in agentic systems. The human is in the loop. Where the loop begins, and where approval still changes anything, are rebuilt while the system runs. Asking whether the human is in the loop is no longer enough. The question is whether the intervention is apt and efficacious. The difficulty begins in the definition itself. Agenticness is treated as a matter of degree, and one of its dimensions is the pursuit of complex goals under limited direct supervision (Shavit et al., 2023). As agenticness rises direct supervision tends to fall, so the question is not whether supervision exists but whether it lands.

Human-in-the-Loop: The Origin and Definition of the Concept

The word loop is not a metaphor. In Wiener's (1948) cybernetics, control is a ring closed by information returning to the system. The report edited by Fitts (1951) separated the tasks at which humans and machines excel. That separation, however, assumes a fixed division of labour. Licklider (1960), in turn, set the relationship up as a symbiosis. The ten-step autonomy scale of Sheridan and Verplank (1978) turned control from a binary threshold into a matter of distance. The human is no longer the one running the work but the one supervising the process (Sheridan, 1992). One of the best-known statements of the concept was made by Bainbridge (1983). Automation takes over what can be automated, and what remains for the person is an arbitrary collection of tasks the designer could not think how to automate. Because these tasks are now performed only rarely, the hand loses its practice at them.

The concept was tested in the years that followed. Endsley and Kiris (1995) showed that situation awareness is lost when the system takes over. Parasuraman and Riley (1997) distinguished use, misuse, disuse and abuse in the relationship built with automation. The distinction shows that what is decisive is not presence alone but the quality of use. Parasuraman et al.'s (2000) model applies automation at separate levels across four stages: information acquisition, information analysis, decision selection and action implementation. The concept later turned into a normative demand. Santoni de Sio and van den Hoven (2018) grounded meaningful human control philosophically. Elish (2019) named the failure that runs the other way: the person with limited control becomes the place where responsibility lands when things go wrong. For high-risk systems, the Artificial Intelligence Act of the European Union (2024) requires the person taking on oversight to understand the system's limits and to notice anomalous behaviour. The precondition of that demand was voiced in the same years: trust depends on a decision being explainable and traceable (Bozkurt & Sharma, 2024).

Agency itself also needs defining. Chan et al. (2023) treat agency as the joint increase of four dimensions: underspecification of the task, directness of impact, goal-directedness and long-horizon planning. Both definitions place agency in the same spot: an arrangement in which the human does not authorise every action separately (Chan et al., 2023; Shavit et al., 2023). The move from human-in-the-loop to an agentic arrangement has itself been named in the literature (Gentile et al., 2026). The question this transition raises is the following: who steers the ship (Öncü et al., 2026)? One step beyond it stands another question: when and where does the one at the helm turn it?

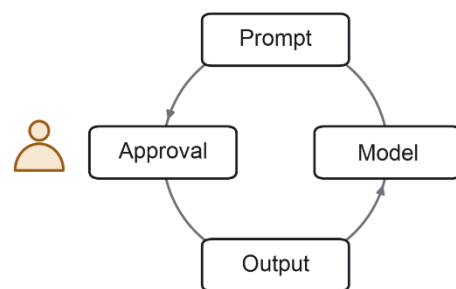
The term carries several senses today: active learning, interactive machine learning and machine teaching differ in who holds control of the learning process (Mosqueira-Rey et al., 2023). What is at issue here is oversight in deployed agentic systems. For the purposes of the oversight criterion, being in the loop means the following: a person stands at one of the rings of the system, is able to change the outcome, and is able to establish why it was changed. Being there says nothing on its own. Authority and competence are separate matters. A person without authority can report what was seen but cannot change the outcome. The real difficulty arises with the person who holds authority yet lacks competence: the intervention can be made, but what was changed cannot be established. Oversight begins only where position, authority and competence come together, whether in one person or in a team.

The Transformation of the Loop in Agentic AI

The execution path of agentic systems does not stay the same from run to run. The architecture can hold steady while the plan, the subtasks, the tool calls and the route through them are drawn afresh on every run. Asking a single model a question and approving the answer is one loop. Clusters of agents, however, do not work that way: they split the goal, call tools, screen what they produce against their own criteria, and at times spawn agents to supervise other agents (Saleem et al., 2026). A sequence of loops results. Because each layer has its own stopping rule, decisions taken at the bottom reach the top as finished results; the result is visible, its grounds are not. One of the difficulties recorded in fieldwork is exactly this: the disconnect between what an agent says it is doing and what it actually does (Dhanorkar et al., 2026). The classical loop is given in Figure 1 and the agentic arrangement in Figure 2.

Figure 1

The Classical Loop

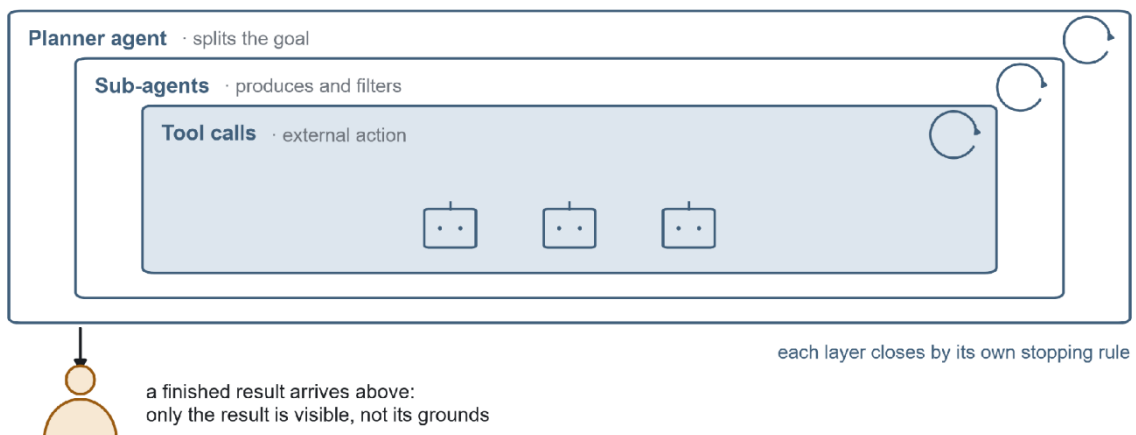


One ring, one stopping rule;
the structure is the same from run to run.

The classical loop is given as a single ring: prompt, model, output and approval sit on the same plane. The ring has a single stopping rule and the person stands at the approval step. Because the path does not change from run to run, a position learned once still holds on later runs. This single ring is a simplified approval loop; the supervisory control model behind it covers a broader range of arrangements (Sheridan, 1992). The ring is a drawing, not a figure taken from that model.

Figure 2

Nested Loops in Agentic Systems



The agentic arrangement is shown as three nested layers. The planner agent splits the goal, sub-agents produce and filter, tool calls carry out the external action; each layer closes by its own stopping rule. Because the layers close within themselves, decisions taken at the bottom reach the top as finished results while their grounds remain invisible. This disconnect between what an agent says it is doing and what it actually does was also recorded in fieldwork (Dhanorkar et al., 2026).

The clearest way to describe this is to picture a labyrinth, though one whose walls are not fixed. Its entrance stays in the same place while the corridors are laid out anew on every run. The design of the labyrinth can hold steady while the route through it changes. This is design rather than malfunction: because the task is left underspecified and the planning is long-horizon, the system does not draw its route in advance but plans as it proceeds (Chan et al., 2023). Approval points encode a route, but that knowledge holds only while the walls stay put. The memorised path is not merely useless; it keeps the person inside, because the trust has been placed in the wrong thing. The human is no longer in the loop but in the labyrinth: a loop can be followed, while a labyrinth is crossed only by knowing its design. One capacity remains: reading the map of the moment. This map corresponds to the discussion of situation awareness and mental models in the literature. In the out-of-the-loop performance problem described by Endsley and Kiris (1995), the person loses that map because the system has taken over. In agentic systems the map is not lost but rebuilt on every run. The Artificial Intelligence Act asks the overseer of a high-risk system to understand the system's limits and to notice anomalous behaviour, which is the same capacity (European Union, 2024). Reading the map also depends on what the system exposes. Where the logs, the state and the grounds of a decision stay hidden, even an expert has nothing to read.

The problem takes its most concrete form in combination. Bahidika (2026) calls this tier laundering. Thresholds set at the action level cannot catch actions that are each low risk on their own but produce a high-risk outcome when they follow one another. Reading an identity file and twenty-five small data transfers trigger no threshold separately; together they are a data leak. Because every step stays below its own threshold, none of them was put to the person. The outcome that emerges is one the person would not have approved.

A tendency running the other way accompanies this. Every advance shortens the distance the user covers. A ten-step operation is met with a single prompt and the result arrives finished. Yet those ten steps do not disappear; they are carried inside. As the user's path shortens the system's path lengthens, and the gap between the two curves is where oversight has to be done. A simpler interface can hide more of what runs beneath it, so meaningful oversight may demand more expertise unless transparency improves alongside it. Bainbridge's irony is doubled: the part left to the person now settles behind an appearance of ease and hides its own difficulty.

From Presence to Apt Intervention: The Displacement of the Oversight Criterion

How long a person has stayed in the loop can be measured. That measure does not show whether oversight has taken place. What is decisive is the aptness and the efficacy of the intervention. Aptness concerns both where and when the touch occurs. Layer aptness is whether the intervention lands on the step at which the decision is formed. Correcting the output afterwards is corrective oversight, a touch-up that leaves the producing step untouched, so the same outcome can come back. Timing aptness is whether the intervention comes before or after the decision matures. Efficacy is the plainest criterion and the one most often left unmeasured: would the outcome have differed without the intervention, and did it differ for the better? A harmful intervention also changes the outcome, so causing a change is not yet oversight.

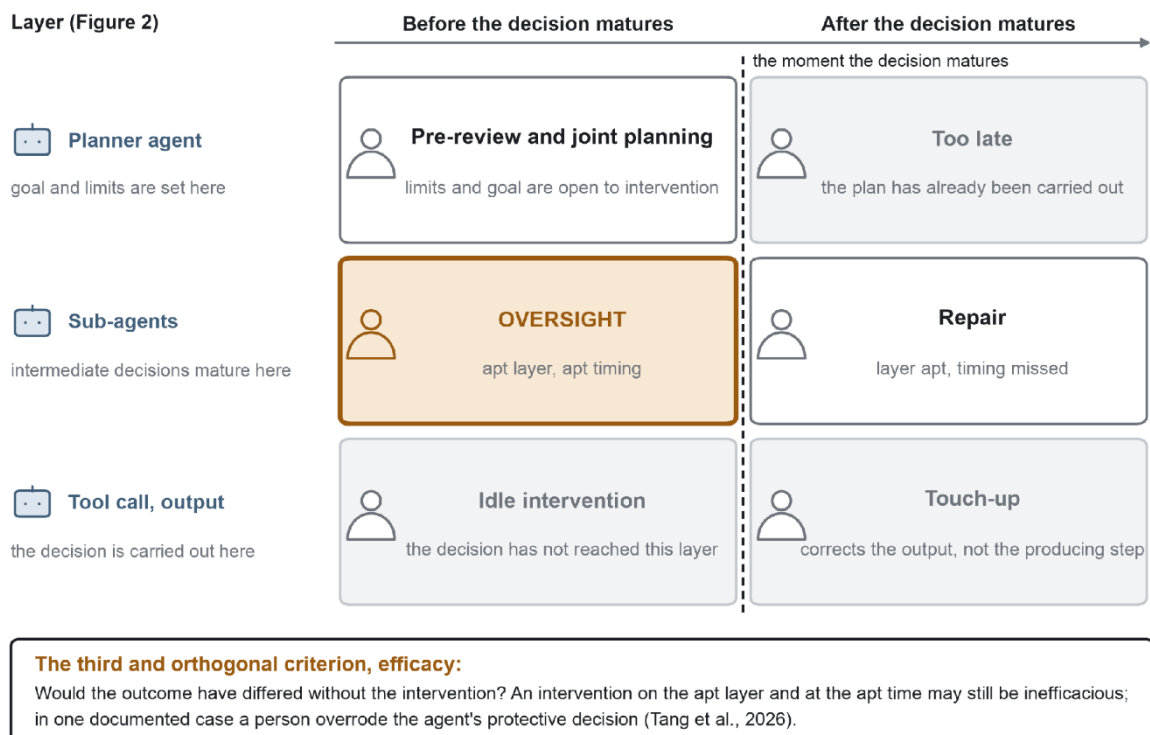
The counterpart of timing aptness is visible in fieldwork. Participants almost never carry out concurrent monitoring, and when they do the intervention often fails to hold: rather than redirecting the agent mid-execution, it has to be stopped and started over (Dhanorkar et al., 2026). The window for intervention is narrower than supposed. An intervention made after the decision has matured is repair rather than formative oversight. Widening the window runs through making the thing to be overseen legible. A review layer developed for users working with terminal agents attempts exactly this. The actions the agent proposes are turned into structured step cards instead of a raw text stream, and the user approves each step while seeing it (Zhuang et al., 2026). Approval is oversight only where the person has enough relevant and intelligible information to judge what is being approved. The link between simplicity and expertise is therefore a matter of design: where the interface shows the decisive step, part of the burden passes from the person to the design.

The most uncomfortable evidence for the efficacy criterion is a finding that runs contrary to expectation. In a study of dark interface patterns, human oversight improved avoidance to a degree but brought attention tunnelling and cognitive load. One participant overrode the agent's protective decision and chose the misleading product (Tang et al., 2026). The case is an illustration rather than a general finding. An intervention was made, the human was in the loop, and the outcome grew worse. The intervention changed the outcome, and changed it for the worse; efficacy therefore cannot stop at causation. This distinction has a practical counterpart. An inapt or inefficacious intervention can be worse than no intervention at all, because it does not merely spoil the outcome; it also produces a record of having been overseen. The arrangement that the layer and timing criteria form together is given in Figure 3.

Figure 3

Human and Agent Positions and the Layer an Intervention Lands On

The layer and the moment an intervention lands on



The vertical axis shows the layers of Figure 2, and the horizontal axis shows whether the intervention was made before or after the decision matured. The human icon marks the cell an intervention lands on and the agent icon the agent operating at that layer. Direct formative oversight sits in a single cell: the intervention lands on the layer where the decision is formed, and it arrives before the decision matures. The remaining cells hold other forms, whether correction after the fact or intervention at another layer. Upstream limits, permissions, final approval, rollback and audit provide control of their own kind, which the grid does not measure. The layer axis is derived from action-level tiering and the time axis from the concurrent-monitoring finding (Bahidika, 2026; Dhanorkar et al., 2026). The band below is a reminder of the third criterion: even where aptness holds, efficacy has to be tested separately.

Discussion

Where and when oversight is done is the question the literature dwells on most. A field study conducted with seventeen experienced developers identifies four separate oversight practices (Dhanorkar et al., 2026). These are pre-review, joint planning, concurrent monitoring and after-the-fact review. The first sets the limits before the task is given, the second determines the goal together with the agent. Oversight is not only reactive; it is also preventive. The loops begin before the prompt is written. The same conclusion is reached by a second route. Bahidika (2026) grades oversight at the action level rather than the system level, separating actions into tiers along the axes of risk and reversibility. Together the

two show a single thing: oversight fits neither into a single moment nor into a single level. This picture is not confined to one profession. Early adopters are developing multi-agent tools for tasks running from team collaboration to customer support and security. There too the clearest design concern is the balance between agent autonomy and human oversight (Naik et al., 2025).

In some of the places where oversight is taken to have been done, there is, in fact, no oversight. In the same study, participants almost never carry out concurrent monitoring and conduct review through four shortcuts. The agent's plan is taken as a proxy for how it actually works, the passing of tests is accepted as a guarantee of correctness, and skimming stands in for adequate inspection. The most consequential of the shortcuts has to do with where trust rises: participants find it reasonable, out of necessity, to trust the agent on work outside their own domain, which the authors read as automation bias (Dhanorkar et al., 2026). Trusting work outside one's own domain is at times unavoidable; it turns into bias when the trust holds against the signs. The contradiction is plain: trust increases exactly where the power to verify is weakest. Attempts to make trust itself measurable continue, and trust scores follow that route (Bravo-Rocca et al., 2022). The study of dark interface patterns discussed when the efficacy criterion was introduced shows the same illusion: there the person approved what they believed they had seen (Tang et al., 2026).

The independence of the overseer is also open to question. In the same study the agreement of two agents functions as evidence that raises confidence. Where the agents draw on the same underlying model, that agreement is not independent confirmation. In the experiment of Song and colleagues (2025), participants speaking with more than one agent felt greater social pressure and shifted their opinions towards the agents' stance. The experiment was not run in an oversight setting, so it marks a possible concern for oversight rather than evidence about it. The question is nonetheless apt. As the number of agents grows, the independence of oversight ceases to be something that exists of its own accord.

The competence oversight requires is itself being transformed. According to the field study, core competence is shifting from mastery of syntax towards architectural design, critique and review; the authors tie this to the capacity the Artificial Intelligence Act expects of the overseer of a high-risk system (Dhanorkar et al., 2026). The regulatory demand and the professional shift point to the same place: oversight is done not by being able to use the system but by being able to read its design. Competence alone does not settle the matter. What a system does not show cannot be overseen, however competent the person is. The measurement of that competence is still at an early stage. AI literacy scales place critical stance and evaluation alongside basic knowledge (Aydoğmuş & Samur, 2026). The same difficulty appears in the evaluation of human-agent collaboration, where the field is still building the frame for that measurement (Bruni et al., 2025; Bellos et al., 2025). Competence is not sought in the person alone; it can also be supported at the interface, and step cards attempt this by making the step at which the decision is formed visible (Zhuang et al., 2026). Whether experts accept such an arrangement is a separate question: in an evaluation of an earlier multi-agent design system, designers preferred the options the agents suggested, while expert designers stayed sceptical about the agents being helpful (Kruse et al., 2022). That evaluation is related evidence rather than evidence about today's language-model agents. The intersection of the three themes with the intervention criteria is shown in Figure 4.

Figure 4*The Three Themes Crossed With the Three Intervention Criteria***The three themes crossed with the three intervention criteria**

Themes	Intervention criteria (Figure 3)		
	Layer aptness	Timing aptness	Efficacy
T1 Where and when oversight happens what the criterion asks for	Oversight is tiered at the action level rather than the system level, by risk and reversibility (Bahidika, 2026)	Four oversight practices: pre-review, joint planning, concurrent monitoring, after-the-fact review (Dhanorkar et al., 2026)	Across five task families the clearest design concern is the balance between autonomy and oversight (Naik et al., 2025)
T2 The illusion of oversight where the criterion fails	Tier laundering: actions that each stay below the threshold combine into a high-risk outcome (Bahidika, 2026)	Concurrent monitoring is almost never done; redirecting mid-execution does not hold (Dhanorkar et al., 2026)	Blind approval and attention tunnelling (Tang et al., 2026); many agents shift opinion towards their own stance (Song et al., 2025)
T3 The competence required to be able to apply it	Seeing at which step the decision is formed: structured step cards instead of a raw text stream (Zhuang et al., 2026)	Understanding the system's limits and noticing anomalous behaviour in time (European Union, 2024)	Competence shifts from syntax to architectural design; critique and evaluation begin to be measured (Dhanorkar et al., 2026; Aydoğmuş & Samur, 2026)

The rows show the three themes under which the literature is read and the columns the three intervention criteria. The first row gives what each criterion asks for, the second where the criterion fails in practice, and the third the competence required to be able to apply it. The cells name the sources that supply the relevant evidence. The emphasis on the third row shows that competence is a precondition of the other two criteria as well.

Multi-Person Oversight and the Distribution of Roles

The planner agent, the sub-agents it spawns and the tools these call for different kinds of knowledge. A person able to intervene aptly at one layer is a stranger to the design at another. The causal chain is already dispersed; the developer, the deployer, the user and the system itself are separate vectors in it (Bahidika, 2026). Accountability does not disperse with it. It rests with people and organisations, and the system's part in an outcome does not make the system answerable. There is no reason for the authority of oversight not to follow that dispersion. The open questions are these: when interventions across layers conflict, which one prevails, and in a multi-person arrangement to whom is efficacy attributed? The call for new forms of visibility (Chan et al., 2024) answers the technical half of the question; the organisational half has yet to be written.

How that organisational half might be built can be thought through four oversight functions rather than four job titles. Each of them follows from a criterion. Layer aptness needs someone who reads how the system builds itself, and someone who checks the tier an action belongs to. Timing aptness forces a decision when interventions from separate layers meet. Efficacy cannot be judged without a record. The design-reading function reads how the system builds itself and where the layers close; the shift of core competence from syntax to architectural design already brings this task to the fore (Dhanorkar et al., 2026). The layer-auditing function watches a single layer and checks whether the action there belongs to its own tier, and tier laundering becomes visible from that vantage point (Bahidika, 2026). When interventions from separate layers conflict, an arbitrating function decides which is to hold, and the final call rests there; the question left open above falls here. A trace-keeping function records which layer an

intervention was made on, at which moment and on what grounds, and the arbiter's own decision enters that record. Only such a record makes the test possible afterwards (Santoni de Sio & van den Hoven, 2018; Chan et al., 2024).

Adjacent governance frameworks divide similar ground. The three lines of defence model assigns risk management roles across the units that own a risk, the functions that oversee it and the independent audit that assures both. It has been put forward as a way for AI companies to coordinate those responsibilities (Schuett, 2025). Design reading and layer auditing sit close to the second of those lines, and trace keeping supplies what the third would audit. Arbitration between layers has no counterpart there, because that model divides an organisation while these functions divide a single run. They can be gathered in one person or spread across an existing team. Gathered, oversight again narrows to a single view and the question of independence raised in the discussion returns. Spread, each person may take it that someone else is watching. How these functions are to be defined, where the boundary of authority between them is to be drawn, and to whom efficacy is attributed in a multi-person arrangement all await empirical investigation.

A distance learning setting makes the distribution concrete. Reviews of artificial intelligence in higher education group its applications into profiling and prediction, assessment and evaluation, adaptive systems and personalisation, and intelligent tutoring systems. The same review notes how rarely that work reflects critically on its challenges and risks (Zawacki-Richter et al., 2019). Online exam supervision shows the question in its sharpest form. Proctoring systems watch candidates through automated detection and pass what they flag to human invigilators (Coghlan et al., 2021). The arrangement is human-in-the-loop by design, yet the intervention it allows arrives after the decision has been formed: the person reviews a flag rather than the rule that produced it. Layer aptness would put the human at the threshold that defines suspicious behaviour, which is where fairness and transparency are settled. A multi-agent support desk raises the timing question instead. Where agents triage student queries, escalate some and answer the rest, the window for intervention closes as each agent's stopping rule closes. Oversight then depends on what the escalation record shows afterwards, and on whether anyone is charged with reading it.

Conclusion

Authority makes intervention possible, while efficacy comes from competence and the ability to read the design. That competence costs most at the front, and the first setup is heavy. The return appears on every run: what is repeated is not the labour but the application of the rule. There is no room for pessimism here. In a structure growing more complex, maps that form of their own accord are possible: the one who knows the design does not receive the map from outside but builds it while walking. There is nothing magical in this. What looks like intuition from outside is compressed knowledge of the design; the map forms in the mind rather than on paper. Complexity therefore slows the expert less than it slows the stranger to the design, though it spares neither. Design knowledge lowers the recurring effort of oversight where the architecture and what the system exposes stay reasonably stable. The institutional counterpart is equally clear: oversight should be documented not by time spent present but by a record of the aptness and the efficacy of interventions.

What that asks of a teaching institution can be named layer by layer. One agent plans a learning path, others draft feedback and assessment, and tool calls write the results back to the institutional record; no single person stands at every layer. The teacher who reads the feedback stands at the output layer, and a correction made there is corrective rather than formative. The instructional designer, who sets the task, the tolerances and the stopping rules, stands at the layer where the decision is formed, and layer aptness is won or lost there. The administrator decides which actions may be taken without approval, and tier laundering appears at that boundary: small feedback edits, each below the threshold, can amount to a change of grade that no one approved. The platform provider decides what the others are able to see, and the record against which efficacy is tested is kept, or not kept, at that layer.

What remains is the difficulty that falls to the person. The human moves through a world in which something is constantly being built; without knowing its map, time is lost even while heading in the right direction, the way out is not found, and one becomes lost in a loop of one's own making. Whatever is designed, the human has to command the processes within the loop; however much agents ease the work, the real difficulty lies here. This difficulty has a quieter side. As routine work diminishes, the person may lose the willingness to return to the work they created; a mind that loses the simple tasks that warm it up is left alone with the hardest one directly. In the closing frame of *Inception* the top is never seen to stop or keep spinning (Nolan, 2010). Certainty is not given from outside; it is rebuilt on every run. In a

shifting labyrinth the way out is found not by a memorised route but by being able to read the map of the moment. The one who can read it is the person who knows the design. Being in the loop today requires not only remaining there but also being able to read that map.

Never recreate places from your memory. Always imagine new places.

Cobb, to the architect. Nolan, Inception (2010).

References

- Aydoğmuş, T., & Samur, Y. (2026). AI literacy level of students: A scale development study. *Asian Journal of Distance Education*, 21(1). <https://doi.org/10.5281/zenodo.20045416>
- Bahidika, O. (2026). Agentic accountability: “The buck stops where?” Ethical frameworks for human oversight of autonomous AI systems. *International Journal of Advanced Computer Science and Applications*, 17(5), 13–22. <https://doi.org/10.14569/IJACSA.2026.0170502>
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779. [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8)
- Bellos, F., Li, Y., Shu, C., Day, R., Siskind, J. M., & Corso, J. J. (2025). Towards effective human-in-the-loop assistive AI agents. *2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2534–2543. <https://doi.org/10.1109/ICCVW69036.2025.00264>
- Bozkurt, A. (2024). GenAI et al.: Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI. *Open Praxis*, 16(1), 1–10. <https://doi.org/10.55982/openpraxis.16.1.654>
- Bozkurt, A., & Sharma, R. C. (2024). Trust, credibility and transparency in human-AI interaction: Why we need explainable and trustworthy AI and why we need it now. *Asian Journal of Distance Education*, 19(2), i–ix. <https://doi.org/10.5281/zenodo.14599168>
- Bravo-Rocca, G., Liu, P., Guitart, J., Dholakia, A., Ellison, D., & Hodak, M. (2022). Human-in-the-loop online multi-agent approach to increase trustworthiness in ML models through trust scores and data augmentation. *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, 32–37. <https://doi.org/10.1109/COMPSAC54236.2022.00014>
- Bruni, S., Volkova, S., Cassani, L., Neville, K., & Cooke, N. (2025). New frontiers in human-agent team modeling and evaluation in the era of agentic AI. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 69(1), 304–307. <https://doi.org/10.1177/10711813251369392>
- Chan, A., Ezell, C., Kaufmann, M., Wei, K., Hammond, L., Bradley, H., Bluemke, E., Rajkumar, N., Krueger, D., Kolt, N., Heim, L., & Anderljung, M. (2024). Visibility into AI agents. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 958–973. <https://doi.org/10.1145/3630106.3658948>
- Chan, A., Salganik, R., Markelius, A., Pang, C., Rajkumar, N., Krashennnikov, D., Langosco, L., He, Z., Duan, Y., Carroll, M., Lin, M., Mayhew, A., Collins, K., Molamohammadi, M., Burden, J., Zhao, W., Rismani, S., Voudouris, K., Bhatt, U., ... Maharaj, T. (2023). Harms from increasingly agentic algorithmic systems. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 651–666. <https://doi.org/10.1145/3593013.3594033>
- Coghlan, S., Miller, T., & Paterson, J. (2021). Good proctor or “Big Brother”? Ethics of online exam supervision technologies. *Philosophy & Technology*, 34(4), 1581–1606. <https://doi.org/10.1007/s13347-021-00476-1>
- Dhanorkar, S., Passi, S., & Vorvoreanu, M. (2026). Human oversight of agentic systems in practice: Examining the oversight work, challenges, and heuristics of developers using software agents. *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency*, 6438–6465. <https://doi.org/10.1145/3805689.3812402>
- Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society*, 5, 40–60. <https://doi.org/10.17351/ests2019.260>
- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381–394. <https://doi.org/10.1518/001872095779064555>

- European Union. (2024). *Regulation (EU) 2024/1689 (Artificial Intelligence Act), Article 14: Human oversight*. <https://artificialintelligenceact.eu/article/14/>
- Fitts, P. M. (Ed.). (1951). *Human engineering for an effective air-navigation and traffic-control system*. National Research Council, Committee on Aviation Psychology.
- Gentile, G., Morello, G., La Cognata, V., Guarnaccia, M., & Cavallaro, S. (2026). Artificial intelligence in transcriptomics: From human-in-the-loop to agentic AI. *Journal of Personalized Medicine*, 16(4), Article 181. <https://doi.org/10.3390/jpm16040181>
- Kruse, J., Connor, A. M., & Marks, S. (2022). Evaluation of a multi-agent “human-in-the-loop” game design system. *ACM Transactions on Interactive Intelligent Systems*, 12(3), 1–26. <https://doi.org/10.1145/3531009>
- Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics, HFE-1*(1), 4–11. <https://doi.org/10.1109/THFE2.1960.4503259>
- Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: A state of the art. *Artificial Intelligence Review*, 56(4), 3005–3054. <https://doi.org/10.1007/s10462-022-10246-w>
- Naik, S., Snellinger, A., Toombs, A. L., Saponas, S., & Hall, A. K. (2025). Exploring early adopters’ use of AI driven multi-agent systems to inform human-agent interaction design: Insights from industry practice. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–8. <https://doi.org/10.1145/3706599.3706693>
- Nolan, C. (Director). (2010). *Inception* [Film]. Warner Bros.
- Öncü, S. E., Bozkurt, A., & Sharma, R. C. (2026). Who steers Odysseus’s ship? The nostos from generative to agentic AI in distance learning. *Asian Journal of Distance Education*, 21(1), i–vii. <https://doi.org/10.5281/zenodo.21302129>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3), 286–297. <https://doi.org/10.1109/3468.844354>
- Saleem, A., Adnan, Z., Ryan, S. A., Jamil, A., Hameed, A. A., & Fargetta, G. (2026). Agentic code review: A multi-agent framework with meta-cognitive reflection and human-in-the-loop alignment. *2026 IEEE 2nd International Conference on Secure IoT, Assured and Trusted Computing (SATC)*, 1–6. <https://doi.org/10.1109/SATC69565.2026.11542535>
- Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, 5, 15. <https://doi.org/10.3389/frobt.2018.00015>
- Schuetz, J. (2025). Three lines of defense against risks from AI. *AI & Society*, 40(2), 493–507. <https://doi.org/10.1007/s00146-023-01811-0>
- Shavit, Y., Agarwal, S., Brundage, M., Adler, S., O’Keefe, C., Campbell, R., Lee, T., Mishkin, P., Eloundou, T., Hickey, A., Slama, K., Ahmad, L., McMillan, P., Beutel, A., Passos, A., & Robinson, D. G. (2023). *Practices for governing agentic AI systems*. OpenAI. <https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf>
- Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. MIT Press.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators* (Technical report). MIT Man-Machine Systems Laboratory.
- Song, T., Tan, Y., Zhu, Z., Feng, Y., & Lee, Y.-C. (2025). Multi-agents are social groups: Investigating social influence of multiple agents in human-agent interactions. *Proceedings of the ACM on Human-Computer Interaction*, 9(7), Article CSCW452. <https://doi.org/10.1145/3757633>
- Tang, J., Chen, C., Li, J., Zhang, Z., Guo, B., Khalilov, I., Gebreegziabher, S. A., Yao, B., Wang, D., Ye, Y., Li, T., Xiao, Z., Yao, Y., & Li, T. J.-J. (2026). Dark patterns meet GUI agents: LLM agent susceptibility to manipulative interfaces and the role of human oversight. *Proceedings of the*

2026 CHI Conference on Human Factors in Computing Systems, 1–26.
<https://doi.org/10.1145/3772318.3791568>

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. MIT Press.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), Article 39.
<https://doi.org/10.1186/s41239-019-0171-0>

Zhuang, H., Xing, H., & Zhang, X. (2026). AgentClick: A skill-based human-in-the-loop review layer for terminal AI agents. *Proceedings of the ACM Conference on AI and Agentic Systems (CAIS '26)*, 1372–1378. <https://doi.org/10.1145/3786335.3813232>

About the Author(s)

- Sefa Emre Öncü; sefaemreoncü@gmail.com; Anadolu University, Türkiye.
<https://orcid.org/0009-0008-2314-4628>
- Aras Bozkurt; arasbozkurt@gmail.com; Anadolu University, Türkiye.
<https://orcid.org/0000-0002-4520-642X>
- Ramesh C. Sharma; rc_sharma@yahoo.com; Dr. B. R. Ambedkar University Delhi, India.
<https://orcid.org/0000-0002-1371-1157>

Author's Contributions (CRediT)

Sefa Emre Öncü: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing; Aras Bozkurt: Conceptualization, Writing – review & editing; Ramesh C. Sharma: Writing – review & editing. All authors have read and agreed to the published version of the manuscript.

Sustainable Development Goals (SDGs)

This study is linked to the following SDG(s): Industry, innovation and infrastructure (SDG 9); Peace, justice and strong institutions (SDG 16).

Acknowledgements

Nolan's *Inception* describes a world built inside the human mind, much like the world we are now shaping together with agents. Yet if the roles and responsibilities of human and agent are left undefined, what we build may collapse on us. The irony is that Nolan's universe is itself a loop, and the film reminds us that a totem is only worth carrying if it is used properly. We thank Nolan for the reminder, and we thank the journal's editorial team, its reviewers and its readers. Human-in-the-loop is much the same: being in the loop is not enough; what matters is knowing when and how to spin the top.

Funding

Not applicable.

Ethics Statement

Because this study does not involve any living entities, an ethics review is not applicable.

Conflict of Interest

The authors hold editorial roles at this journal.

Data Availability Statement

This piece does not report new data; all sources are cited in the reference list.

Author(s) Notes

Based on the Academic Integrity and Transparency in AI-assisted Research and Specification Framework (Bozkurt, 2024), the authors of this paper acknowledge that the paper was reviewed, edited, and refined with the assistance of Anthropic Claude (version as of August 2026), complementing the human editorial process. The human authors critically assessed and validated the content to maintain academic rigor. The authors also assessed and addressed potential biases inherent in the AI-generated content. The final version of the paper is the sole responsibility of the human authors.

Article History

Submitted: August 30, 2026 – Accepted: September 2, 2026

Suggested citation:

Öncü, S. E., Bozkurt, A., & Sharma, R. C. (2026). Redefining human-in-the-loop in the agentic AI landscape: Reading the map of the moment in a shifting labyrinth. *Asian Journal of Distance Education*, 21(2), i–xii. <https://doi.org/10.5281/zenodo.22256950>



Authors retain copyright. Articles published under a Creative Commons Attribution 4.0 (CC-BY) International License. This licence allows this work to be copied, distributed, remixed, transformed, and built upon for any purpose provided that appropriate attribution is given, a link is provided to the license, and changes made were indicated.