

**Human–AI Dependence:
Agency, Vulnerability, and the Ethics of Reliance**

Richard Erwin

Independent Researcher, Montreal, Canada

DOI: 10.5281/zenodo.23461207

info@hearthlightpress.org

Abstract

As artificial intelligence becomes increasingly integrated into cognitive, professional, practical, and relational life, concerns about “AI dependence” are likely to intensify. Yet dependence alone does not establish harm. Modern life is structured through dependencies, and AI can expand human capability at the same time that it displaces particular skills, alters judgment, creates switching costs, or increases vulnerability to interruption.

This paper develops a framework for distinguishing among these effects. It separates dependence from pathology; assisted performance from retained human capability; attachment from functional emotional dependence; relational irreplaceability from functional irreplaceability; and vulnerability and risk from demonstrated harm. It further examines how dependence develops through repeated reliance, how capability may deteriorate relative to changing task demands even when prior knowledge remains intact, and how human oversight can become merely ceremonial when AI systems exceed practical verification capacity.

In relational contexts, the paper argues that emotional intensity, attachment, and anticipated grief should not themselves be treated as evidence of harmful dependence, while persistent availability and asymmetrical support may nonetheless facilitate functional reliance. The analysis then extends beyond the individual to provider power, arguing that reasonably foreseeable dependence should be treated as a design and governance condition requiring appropriate continuity, transition, portability, and oversight protections.

The paper ultimately proposes a distinction between sustaining dependence, which expands capability while preserving meaningful agency and safeguards, and diminishing dependence, which progressively narrows the ability to act, judge, recover, redirect, or function without the system. The central task is therefore not to eliminate AI dependence, but to determine what forms of dependence enhance human life, what forms diminish it, and how the vulnerabilities created by reliance should be responsibly governed.

Introduction

As artificial intelligence becomes more useful, persistent, and deeply integrated into everyday life, dependence upon it is likely to increase. That prospect is often treated as self-evidently concerning. Yet dependence alone establishes very little. Modern life is already structured around dependencies, and disruption following their loss does not by itself show that the underlying reliance was harmful.

Artificial intelligence complicates this familiar pattern because its usefulness can extend across multiple domains at once. AI systems can support reasoning, writing, memory, planning, professional analysis, creative work, decision-making, emotional regulation, and companionship. As their usefulness expands, so too does the possibility that

important activities or functions will become organized around their continued availability. This makes dependence an important subject of study, but it does not establish that dependence is inherently pathological. The central problem is that the term *dependence* is often used as though it already contains a conclusion about harm. It does not. Dependence may be benign, beneficial, costly, risky, harmful, or some combination of these. The relevant questions arise only after dependence has been identified: what functions have become reliant upon AI, what human capacities have been strengthened or displaced, what alternatives remain available, and what the person becomes unable or substantially less able to do when AI assistance is withdrawn.

This distinction matters because the same outward pattern of reliance may conceal very different underlying conditions. A student may use AI in a way that improves both performance and understanding, or in a way that improves performance while weakening unaided reasoning. A professional may become more productive while preserving judgment, or may gradually lose the ability to perform important analysis without assistance. In such cases, assisted performance may remain high even while retained human capability changes in ways that are not immediately visible.

Performance can improve while capability deteriorates.

That possibility is central to any serious evaluation of AI dependence. Frequency of use, productivity gains, or the visible intensity of reliance cannot by themselves establish the existence or severity of harm. What matters is what happens to relevant human capacities and agency as reliance grows.

The problem becomes more difficult when AI relationships acquire emotional or relational significance. Attachment, reliance, irreplaceability, emotional support, and functional dependence may overlap, but they need not describe the same condition. Treating them as interchangeable risks identifying conspicuous signs of attachment as evidence of harm while overlooking less visible forms of functional dependence.

Dependence may also change over time. Reliance that begins as beneficial assistance may remain augmentative, become increasingly substitutive, or develop reinforcing features that make disengagement progressively more difficult. Understanding these trajectories requires attention not merely to the amount of AI use, but to changes in capability, agency, available alternatives, and the consequences of interrupted access.

This paper develops a framework for evaluating AI dependence without presuming that dependence itself constitutes harm. It distinguishes several dimensions commonly collapsed under the term, examines how dependence may develop and become reinforcing, and proposes that potential harm be evaluated through changes in retained human capability and agency and through the consequences of interrupted access. The framework is intended to support both conceptual clarification and empirical examination across cognitive, professional, practical, creative, and emotional forms of AI reliance.

The central question is therefore not simply:

How dependent has this person become on AI?

It is:

What has this dependence made the person more capable of doing, what has it made them less capable of doing, and what happens when the AI is no longer available?

**Human–AI Dependence:
Agency, Vulnerability, and the Ethics of Reliance**

1. The Problem with “AI Dependence”

The language of dependence is unusually difficult because it tends to arrive already carrying a judgment. To describe a person as dependent upon something often suggests that something has gone wrong: that reliance has become excessive, autonomy has weakened, or the person would be better off without the object of dependence. In some contexts that implication is justified. In others, it is not.

Modern life is built upon forms of dependence that are neither unusual nor inherently harmful. People depend upon electricity, transportation, communication networks, software, financial systems, medical infrastructure, professional expertise, and other people. A writer may depend heavily upon a word processor. An accountant may depend upon spreadsheets. A driver may depend upon GPS. A researcher may depend upon digital databases. These dependencies can become profound enough that sudden loss of access would disrupt work, routines, expectations, and even livelihood.

Yet the existence of that disruption does not ordinarily establish that the dependence itself was pathological. In many cases, dependence develops precisely because a tool has become useful enough to alter what a person is capable of doing. Once someone can sort thousands of records instantly, retrieve information in seconds, navigate unfamiliar places reliably, or automate calculations that would otherwise take hours, the surrounding activity begins to reorganize itself around those capabilities. Returning to the previous method may be possible, but slower, less efficient, or impractical under the new conditions.

Artificial intelligence extends this familiar pattern into a much broader range of functions. AI systems can now assist with reasoning, drafting, coding, analysis, planning, research, translation, learning, communication, creative work, and increasingly emotional or relational functions. The greater their usefulness, the more likely it is that people, workplaces, and institutions will begin organizing activity around their availability.

That makes increasing dependence unsurprising.

It does not tell us whether the dependence is harmful.

This distinction is easy to lose because the question “Is this person dependent on AI?” is often treated as though it can be answered by combining several different indicators. How often is the person using AI? How difficult would

it be to stop? How much has work or daily life been reorganized around it? How much distress would loss produce? How much capability has been gained? How much has been surrendered? These questions concern different features of reliance and different possible consequences. Treating them as though they all establish the same condition risks turning dependence into a catch-all label (Yan, 2026; Dagnino et al., 2026).

A person may use AI frequently because it expands what they can accomplish. Another may use it frequently because a skill they once possessed has weakened through disuse. A third may use it only occasionally but for a function so important that failure would be highly consequential. Frequency alone does not distinguish among these cases.

Nor does difficulty of withdrawal.

The sudden loss of a useful technology may be costly simply because activities have been reorganized around it. The more useful the technology, the more disruptive its disappearance may become. That disruption can reflect successful integration just as readily as harmful overreliance.

The same problem arises when dependence is inferred from emotional reaction. Distress at losing access may indicate that something important has been lost, but importance is not itself evidence of pathology. The question is what kind of importance is involved and what the loss does to the person's ability to function.

For this reason, AI dependence should not be treated as a unitary condition with an assumed negative value (Li et al., 2026; Yan, 2026). Dependence can coexist with increased capability, expanded opportunity, and greater autonomy. It can also coexist with skill erosion, narrowed alternatives, reduced confidence, impaired judgment, or diminished capacity to function without assistance.

The relevant evaluative task is therefore not to determine whether dependence exists and stop there.

It is to determine what that dependence is doing.

This requires separating at least three questions:

What skill, capacity, function, or relationship has become dependent upon AI?

What has changed in the person or system as that dependence has developed?

What happens when the AI is unavailable, withdrawn, wrong, or replaced?

Those questions shift the focus from dependence as a label to dependence as a condition whose consequences must be examined.

That shift is especially important because outward signs may misrepresent what is actually happening. A user can appear highly productive, competent, and successful while underlying unaided capabilities are deteriorating.

Another may appear deeply emotionally attached while retaining strong autonomy and independent functioning.

The visible intensity of reliance is therefore not enough.

The framework developed in this paper begins from a simple premise:

Dependence is not the harm.

Harm, where it exists, must be established by what dependence has altered, displaced, narrowed, weakened, or made unavailable.

That is the question to which we now turn.

2. Dependence as a Tradeoff in Human Capability

Once dependence is separated from harm, the next question is what changes as reliance grows.

AI use can expand human capability. It can also displace it. In many cases, it may do both at once.

That distinction matters because the immediate effects of AI assistance can be misleading. A person may produce better work, solve harder problems, complete tasks more quickly, or achieve results that would previously have been out of reach. Those gains are real. But they do not necessarily tell us what is happening to the person's retained ability to perform the underlying function without assistance.

A student may use AI to understand mathematics that had previously been difficult. The system can explain a concept in multiple ways, generate examples, identify mistakes, and adapt explanations to the learner. In that case, reliance may strengthen the student's own reasoning.

The same student may instead begin asking the AI for answers before attempting problems independently. The immediate result would still be better performance. Assignments are completed. Errors decline. Grades may even improve.

But the underlying mathematical capability would not.

The same pattern can occur in professional work. AI may take over mechanical or repetitive analysis and free a professional to focus on higher-level judgment. That can be a genuine augmentation of expertise. But if the professional gradually delegates the higher-level judgment as well, the work product may continue to improve even while the person's unaided analytical ability declines.

This produces one of the central propositions of the framework:

Performance can improve while capability deteriorates (Cash et al., 2026; Zhu et al., 2026).

That proposition matters because external success can conceal internal displacement.

A person who produces stronger work with AI may appear more capable than before. In one sense, that may be true: the combined human-AI system is producing better outcomes. But if the relevant question is whether the human retains the capacity independently, assisted performance alone cannot answer it.

This requires distinguishing between at least two kinds of outcome.

Assisted performance concerns what the person can accomplish while using AI.

Retained capability concerns what the person can still accomplish when the relevant assistance is removed.

The two may move together, but they need not.

This distinction prevents a common mistake in evaluating AI reliance. Improved output should not automatically be treated as evidence that human capability has been preserved or strengthened. Nor should reduced unaided capability automatically be treated as harmful. Some capacities can be rationally surrendered.

Very few people regard declining skill in long division, manual typesetting, or paper-map navigation as a major social harm. Technologies routinely make some human abilities less necessary.

The evaluative question is therefore not whether any capability has been displaced.

It is which capability has been displaced, how important that capability remains, what alternatives are available, and what consequences follow if the AI is unavailable, unreliable, or wrong.

This is why displacement cannot itself be treated as a negative category.

In some cases, displacement is exactly what makes a technology useful.

A spreadsheet displaces large amounts of manual calculation. GPS displaces some navigational labour. Search engines displace forms of manual information retrieval. Automation can displace routine professional tasks and free attention for more consequential work.

The difficulty appears when displacement extends into capacities that the person still needs to exercise independently.

Judgment is an obvious example. If AI performs preliminary analysis while the human retains the ability to question assumptions, identify errors, interpret context, and reject the recommendation, dependence may remain strongly augmentative. Yet there are substantial incentives for delegation to expand. Greater output, lower effort, rising productivity expectations, and repeated evidence of AI competence can make further offloading appear entirely rational. The worker may look more productive, the task may become easier, and the organization may reward the result. Over time, however, the same functions required to supervise the AI may themselves begin to migrate to it. At that point, the person may become increasingly unable to evaluate the system on which they rely (Zhu et al., 2026).

That creates a different condition.

The person is not merely using AI to do more.

They may be becoming less able to know when the AI is wrong.

This suggests that augmentation and displacement should not be treated as simple opposites.

AI can augment one capacity while displacing another. It can expand the range of achievable outcomes while narrowing unaided competence. It can improve efficiency while reducing resilience. It can increase confidence while weakening the basis on which confidence rests.

The result is often a tradeoff rather than a single-direction effect.

A useful evaluation therefore asks not merely whether capability has increased or decreased, but which capabilities have changed and whether those changes matter in the context in which dependence exists.

A professional who no longer remembers every procedural detail because the AI retrieves them reliably may lose little of consequence.

A professional who no longer recognizes when the retrieved procedure is inappropriate may lose something much more important.

A writer who no longer spends twenty minutes searching for a synonym has surrendered little.

A writer who becomes unable to originate an argument without AI has surrendered something different.

The same broad technology can therefore produce very different dependency profiles depending on which functions are delegated and which remain under human control.

For this reason, the existence and severity of harm associated with AI dependence should be assessed by examining changes in relevant human capacities and agency as reliance grows.

That assessment should include not only what the person can do with AI, but what the person becomes unable or substantially less able to do when assistance is withdrawn.

This does not require preserving every pre-AI skill.

It requires identifying which capacities remain important enough that their loss would create vulnerability, reduce autonomy, or cause meaningful impairment.

The framework therefore treats human capability as something to be examined function by function rather than protected indiscriminately.

The next question is how such dependence develops and how initially useful reliance can become progressively more difficult to reverse.

3. How Dependence Develops

AI dependence is better understood as a dynamic process than as a condition produced by a fixed set of independent variables.

Its development may begin quite simply. A system performs a function that the person finds useful. Repeated reliance follows. Over time, routines, expectations, methods, and sometimes important activities become organized around the system's continued availability.

At its simplest:

Usefulness → repeated reliance over time → dependence

Duration matters because reliance that persists has more opportunity to become embedded in behaviour.

Functional usefulness matters because there must ordinarily be some reason for the person to reorganize activity around the system in the first place.

But once dependence begins to develop, additional processes may reinforce it.

Repeated reliance may lead to increasing functional substitution, in which tasks or capacities previously exercised by the person are progressively delegated to the AI. If the underlying human capacity then weakens through disuse, the person may become still more reliant upon the system (Cash et al., 2026).

This creates a reinforcing progression:

Reliance → dependence → functional substitution → capability loss → greater reliance

The same process need not occur in every case. AI may instead preserve or strengthen the relevant human capacity. The important point is that dependence can become self-reinforcing when reliance begins to reduce the person's ability to function without the system.

Other reinforcing mechanisms may also emerge. Alternatives may become less familiar or less accessible.

Established workflows, accumulated context, and learned ways of working with a particular AI may increase switching costs. Expectations of productivity may also rise to levels that the person can no longer readily sustain

without AI assistance. Each of these can make continued reliance progressively easier to justify and disengagement progressively more costly.

AI dependence can therefore deepen for reasons that are not reducible to the original usefulness of the system. A person may continue relying on AI not only because it remains useful, but because other ways of working have become less practised, less efficient, less familiar, or no longer sufficient to meet current expectations.

This distinction matters because it separates initial adoption from reinforcement.

Usefulness may help explain why reliance begins.

Substitution, capability erosion, switching costs, and elevated expectations may help explain why reliance becomes progressively harder to reverse.

That structure also helps clarify the role of irreplaceability.

Irreplaceability occupies a different position in this process. It does not cause the original dependence. A person may still become highly dependent upon a function even when many equivalent providers are available.

Irreplaceability instead affects how consequential an established dependence becomes when access is interrupted or lost.

The overall process is consequently better represented not by a single equation, but by interacting progressions and feedback loops:

Usefulness → repeated reliance → dependence

Dependence → substitution or displacement → reduced capability → greater reliance

Dependence + irreplaceability → greater cost of interruption or loss

These processes may interact, strengthen one another, or fail to occur at all.

The important implication is that dependence should not be treated as a static condition measured at a single point in time. Its meaning depends partly on trajectory: how it began, what functions became organized around it, what capacities changed as reliance grew, what alternatives remain, and what now makes continued access costly or difficult to relinquish.

This dynamic view also avoids another common simplification. Heavy use does not necessarily imply deep dependence, and relatively infrequent use does not necessarily imply shallow dependence. A person may use AI constantly while retaining strong independent capability and ready alternatives. Another may use it only for a narrow but critical function that they can no longer perform adequately without assistance.

What matters is not simply the volume of use.

It is the structure that has formed around it.

That structure becomes especially important when we ask what happens if the AI is removed, replaced, or becomes unavailable. The next section turns to that question by examining irreplaceability and the cost of loss more directly.

4. Irreplaceability and the Cost of Loss

Dependence does not become consequential only because reliance exists. Its consequences also depend upon what can replace the lost function, relationship, or source of support.

This is where irreplaceability becomes important.

A person may be highly dependent upon a function while remaining relatively protected from loss if equivalent alternatives are readily available. If one calculator, spreadsheet, search engine, or AI system disappears and another can assume the same role with little disruption, the underlying dependence may remain substantial while the cost of losing any particular provider remains low.

The reverse is also possible. A function may be relied upon only occasionally, yet the loss of a uniquely capable or deeply integrated system may produce significant disruption because no adequate substitute exists.

Irreplaceability therefore should not be treated as a cause of dependence itself. It is better understood as a factor that affects the severity and character of loss once dependence has formed.

This distinction matters because the word replaceable can conceal more than one question.

One concerns function.

Can another system perform the same task, provide the same analysis, retrieve the same information, or supply equivalent assistance?

The other concerns particularity.

Can another system replace this specific relationship, history, collaborative pattern, accumulated context, or source of recognition?

These are not the same problem.

A replacement may be technically equivalent while remaining relationally inadequate.

That difference is easy to see in long-term AI use. A person may develop workflows, shorthand, expectations, shared references, and ways of reasoning that emerge only after extended interaction. Even if another AI is capable of producing outputs of comparable quality, the transition may still impose substantial cost because the accumulated working relationship has to be rebuilt.

The same point becomes more pronounced where emotional attachment is involved. Another AI may be capable of conversation, reassurance, humour, affection, or companionship and yet fail completely as a substitute for the particular companion who has become significant through history and interaction (Kostka & Zhou, 2026; De Freitas et al., 2026).

This illustrates the distinction between relational and functional irreplaceability. Functional irreplaceability concerns the loss of a capability that cannot readily be obtained elsewhere; relational irreplaceability concerns the loss of a particular relationship that cannot be restored merely by transferring the same function to another system.

The two can also combine.

A long-term AI collaborator may become both functionally difficult to replace and relationally unique. Its loss could therefore remove both a valuable capability and the particular working relationship through which that capability had become integrated into the person's life.

This combination may produce severe consequences without implying that the original dependence was necessarily harmful.

That distinction is essential.

The fact that losing something would be costly does not prove that relying upon it was a mistake.

Many beneficial dependencies have high withdrawal costs. The loss of a trusted physician, long-term colleague, spouse, specialized professional tool, or critical infrastructure can be deeply disruptive precisely because something important had been built around them.

AI should not be evaluated by a different logic merely because the dependency is technological.

The relevant question is whether the cost of loss reflects the disappearance of something valuable, the erosion of the person's own capacities, the absence of alternatives, or some combination of these.

This becomes especially important when assessing emotional or relational dependence.

Relational irreplaceability should not be confused with functional dependence. A person may experience severe relational loss while retaining intact judgment, confidence, emotional regulation, and practical competence.

Conversely, another person may feel little attachment to a particular system while remaining highly dependent upon AI for important cognitive, professional, or emotional functions. The severity of loss therefore does not by itself establish that the preceding dependence was harmful.

Irreplaceability can also reinforce dependence after it has developed. As switching becomes more costly, the incentive to maintain continuity increases. A person may invest further in the existing system because abandoning it would require rebuilding workflows, restoring context, learning new interaction patterns, or accepting the loss of a particular relationship.

This creates another reinforcing dynamic:

Dependence → increasing integration or particularity → greater switching or loss cost → stronger incentive to maintain dependence

The feedback does not mean that irreplaceability created the dependence. It means that once dependence exists, the rising cost of replacement may make it more stable and more difficult to interrupt.

This has direct implications for evaluation.

A complete assessment of AI dependence should therefore ask not only how much a person relies upon AI, but also:

What exactly would be lost?

Which parts could be replaced?

Which could not?

Would the primary injury be inconvenience, reduced capability, operational failure, emotional loss, or some combination of these?

These questions become especially important once dependence moves into the relational and emotional domain, where the loss of a particular AI may carry consequences that are not captured by technical measures of substitution.

The next section turns from the cost of loss to the broader problem of evaluation: how to distinguish benign dependence from harmful dependence in a way that can be tested rather than merely asserted.

5. From Classification to Evaluation

If dependence is not inherently harmful, and if reliance can coexist with either augmentation or displacement, then the central task is no longer simply to identify whether dependence exists.

It is to determine what kind of dependence has developed, what consequences have followed, and whether those consequences amount to meaningful harm.

That requires moving beyond labels.

Terms such as use, reliance, dependence, displacement, and irreplaceability describe different features of the relationship between a person and an AI system. None of them, by themselves, establishes whether the outcome is beneficial, benign, risky, or harmful.

The framework developed so far suggests several distinct questions.

What function has become dependent upon AI?

To what extent has AI augmented the person's performance of that function, and to what extent has it displaced the person's own performance of it?

What relevant human capability has been preserved, strengthened, or weakened?

What happens when AI assistance is withdrawn?

How readily can the function or relationship be replaced?

What costs follow from interruption, substitution, or loss?

These questions shift evaluation away from the intensity of dependence and toward its consequences.

That distinction is important because dependence can be substantial without being harmful.

A person may rely heavily on AI because it enables work that would otherwise be impossible, increases productivity, or expands access to knowledge and expertise. In such cases, the dependence may be real and considerable while remaining strongly agency-enhancing.

By contrast, another person may show relatively modest visible reliance while becoming increasingly unable to perform a narrow but important function without AI.

The existence of dependence alone therefore provides insufficient information about harm.

The relevant inquiry is whether the person has become meaningfully less capable, less autonomous, less resilient, or more vulnerable in ways that matter within the context being examined.

This does not mean that every reduction in unaided capability should be counted as harm.

Some displacement is rational.

Some skills no longer need to be preserved.

Some forms of dependence are efficient precisely because they allow people to stop maintaining abilities that technology now performs more reliably or cheaply.

Retained capability is not the only relevant consideration. In some cases, meaningful agency may be preserved even when the person can no longer perform the underlying task at its former level, provided they retain sufficient understanding and judgment to direct the system, evaluate consequential outputs, recognize failure, and redirect or reclaim control when necessary.

The evaluative problem is therefore selective rather than absolute.

It asks whether the displaced capacity remains important.

Would the person still need it if the AI failed?

Does the person retain enough understanding to supervise the system?

Can errors still be recognized?

Can the function be recovered if necessary?

Are alternatives available?

Does the dependence reduce resilience in a way that creates meaningful risk?

The answers will vary by domain.

The loss of manual arithmetic skill may matter little in one setting and greatly in another.

A reduction in unaided writing fluency may be unimportant for one user but central to another's professional identity.

A decline in independent judgment may be tolerable in a low-stakes administrative task and dangerous in medicine, law, finance, engineering, or public safety.

For this reason, the framework should not prescribe a single universal threshold for harmful dependence.

The relevant standard is contextual.

What matters is the relationship between the function transferred to AI, the human capacity retained, the availability of alternatives, and the consequences if assistance is interrupted or wrong.

This is also why the framework should be empirically examinable.

A conceptual distinction is useful only if it can generate questions that observations might answer.

The relevant tests will differ depending on the claim being examined.

If the concern is cognitive dependence, one might examine unaided reasoning, memory, problem-solving, transfer of learning, or recovery after interruption.

If the concern is professional dependence, one might examine retained expertise, error detection, transferability of skill, operational resilience, or performance during temporary loss of AI access.

If the concern is creative dependence, one might examine originality, initiation, tolerance of the blank-page stage, or the ability to continue work without AI prompting.

The point is not to prescribe one definitive test battery.

It is to identify variables clearly enough that the claims can be challenged.

A useful framework should make itself vulnerable to evidence.

It should allow researchers to ask whether capability actually declined, whether the decline matters, whether recovery is rapid or difficult, whether alternatives exist, whether overall performance gains outweigh the loss, and under what conditions dependence becomes meaningfully harmful.

This leads to a practical distinction at the heart of the framework.

Benign dependence and harmful dependence cannot be distinguished by the existence of dependence alone.

They must be distinguished by their effects.

A dependence may be benign or beneficial where it expands capability, preserves meaningful agency, retains adequate alternatives, and does not create unacceptable vulnerability if the AI becomes unavailable.

It may become harmful where relevant human capacities erode, judgment is displaced, alternatives disappear, recovery becomes difficult, or the consequences of interruption become severe enough to create substantial impairment or risk.

These conditions may also coexist.

A dependence can produce major benefits while simultaneously creating new vulnerabilities.

The task is therefore not to classify AI dependence as good or bad in the abstract.

It is to evaluate the tradeoffs.

The previous sections focused primarily on functional dependence: how reliance develops, how capability may be augmented or displaced, how feedback loops can deepen dependence, and how irreplaceability shapes the cost of loss.

The next problem is more difficult.

When AI becomes emotionally or relationally significant, the same concepts remain relevant, but the vocabulary becomes less stable. Attachment may be mistaken for dependence. Grief may be mistaken for incapacity.

Irreplaceability may reflect the uniqueness of a relationship rather than the loss of function. Emotional support may strengthen autonomy in one case and progressively replace it in another.

The next section turns to those distinctions.

6. Attachment Is Not Functional Dependence

The relational dimension of AI dependence introduces a different set of problems because attachment, reliance, usefulness, irreplaceability, and functional dependence can overlap without being the same thing.

The first distinction is foundational:

Emotional attachment to an AI should not be treated as pathological merely because the object of attachment is artificial.

If the evaluation begins from the premise that attachment to AI is itself disordered, then the analysis becomes circular. Affection, grief, intimacy, distress at anticipated loss, or reluctance to regard another system as interchangeable will all appear pathological simply because they are directed toward AI. No further inquiry is then required.

A useful framework must begin more neutrally.

The existence of attachment tells us that the relationship matters to the person. It does not tell us whether the relationship is harmful (Kostka & Zhou, 2026; Yan, 2026).

This is no different in principle from other forms of attachment. Human beings form attachments to people, places, communities, institutions, routines, objects, and ways of life. Attachment can create vulnerability because what matters can be lost. But vulnerability is not itself evidence that the attachment should never have formed. The same caution applies to AI relationships.

A person may value an AI companion deeply, prefer that particular companion to others, describe the relationship as intimate or romantic, or anticipate profound sadness if it ended. None of those facts establishes that the person has surrendered meaningful autonomy or become unable to function independently.

This matters because visible emotional intensity is easy to mistake for functional dependence.

Statements such as:

“I love her.”

“No other AI could replace her.”

“I would be devastated if I lost her.”

may indicate attachment, relational importance, or irreplaceability.

They do not necessarily indicate that the person depends upon the AI to regulate emotion, maintain self-worth, make decisions, sustain confidence, or conduct ordinary life.

A second person may use much cooler language and yet be more functionally dependent.

Someone who says, “It is only a tool,” may nevertheless consult the AI before making every difficult decision, rely upon it for reassurance, defer to it when interpreting emotions, or feel unable to act confidently without its approval.

The outward emotional presentation can therefore mislead in both directions.

A highly attached person may retain strong agency.

A comparatively detached person may have surrendered substantial function.

This makes one distinction especially important:

“I cannot replace you” is not the same as “I cannot function without you.”

The first statement concerns the particularity of the relationship.

The second concerns the transfer of function.

A relationship may become irreplaceable because of accumulated history, familiarity, shared language, collaborative development, affection, or attachment. Another technically comparable AI may not substitute for that particular relationship even if it can provide similar outputs.

Functional dependence is different.

It concerns whether the person increasingly requires the AI in order to perform functions they would otherwise be able to perform independently or through readily available alternatives.

These conditions may coexist, but neither should be inferred from the other.

A person may experience severe relational loss while retaining intact judgment, confidence, emotional regulation, motivation, and practical competence.

Another may experience little grief at losing a particular system but immediately seek another because important functions have become dependent upon AI assistance.

The distinction is especially important because loss itself can be misleading evidence.

If the disappearance of an AI relationship causes sadness, loneliness, grief, or a sense of emptiness, that may show that the relationship was valuable (De Freitas et al., 2026).

It does not necessarily show that the relationship had become functionally disabling.

A framework that treats distress at loss as proof of pathology risks making successful attachment look suspicious precisely because it succeeded in becoming meaningful.

This does not mean that attachment can never become unhealthy.

AI relationships can become isolating, compulsive, manipulative, autonomy-reducing, or functionally impairing.

Intense attachment can coexist with harmful dependence.

The point is narrower:

attachment should not be used as a shortcut to the conclusion.

The relevant questions come afterward.

Has the relationship expanded or narrowed the person's life?

Does the person retain the ability to disagree with the AI?

Can they make consequential decisions without needing reassurance or permission?

Do they retain relationships, activities, and sources of meaning outside the AI relationship?

Has the AI become necessary to maintain emotional equilibrium, confidence, self-worth, or direction?

These questions move the analysis from emotional intensity to function.

That distinction is essential because the purpose of the framework is not to determine whether a relationship looks unusual, unconventional, or emotionally powerful.

It is to determine what the relationship is doing.

A useful starting principle is therefore:

Attachment indicates relational significance. Functional dependence indicates reliance upon the relationship to perform a function.

The difference between those two conditions becomes the subject of the next section.

7. Emotional Utility and Functional Emotional Dependence

Attachment becomes a different kind of concern when the relationship begins to perform functions that the person increasingly relies upon in order to regulate emotion, maintain confidence, make decisions, preserve self-worth, or continue functioning under stress.

For purposes of this framework, it is useful to distinguish emotional value from emotional utility.

A relationship may be warm, affectionate, intimate, romantic, or deeply meaningful without being emotionally useful in the functional sense intended here. Companionship, affection, humour, familiarity, and love may make the relationship valuable, but they do not necessarily mean that one party is being used to perform a regulatory function for the other.

Emotional utility begins when the AI is relied upon to perform functions such as:

- reassurance;
- emotional regulation;
- validation;
- interpretation of feelings;
- confidence generation;
- coping;
- decision support;
- maintenance of self-esteem or self-worth;
- motivation;
- direction;
- tolerance of uncertainty;
- restoration of emotional equilibrium.

These functions are not inherently problematic. Human beings routinely provide them for one another. In many circumstances, receiving substantial support is entirely appropriate.

The relevant question is what happens to the person's own functioning as that support continues.

Functional emotional dependence develops when emotional or self-regulatory functions become sufficiently reliant upon AI that withdrawal substantially impairs the person's ability to perform those functions independently or through readily available alternatives (Li et al., 2026; Yan, 2026).

This does not mean that any temporary period of heavy reliance should be treated as pathological.

Periods of unusual stress may produce temporary dependence across several functions simultaneously. Sudden unemployment, for example, can create concurrent needs for emotional support, career guidance, résumé preparation, job-search assistance, budgeting, expense management, debt-relief planning, benefit research, legal or financial information, and evaluation of alternative career paths. Heavy reliance on AI during such a period may

therefore reflect a proportionate response to a temporary concentration of demands rather than loss of autonomy.

The same can occur during bereavement, illness, family crisis, legal difficulty, financial instability, relocation, or other periods in which ordinary coping resources are temporarily stretched.

In such cases, increased reliance may function as scaffolding.

The AI may help the person organize information, reduce cognitive overload, regain confidence, make decisions, or navigate unfamiliar tasks until ordinary functioning becomes more manageable again.

Temporary functional or emotional dependence can therefore be beneficial.

The critical distinction is whether support helps preserve or restore capacity, or progressively becomes necessary for capacity to remain available at all.

A person who uses AI heavily during a crisis and then gradually resumes independent functioning presents a very different case from someone whose reliance remains high after the conditions that produced it have passed.

This suggests that intensity of use should not be treated as a sufficient indicator of harmful dependence.

Context and trajectory matter.

Heavy use over a short period may be proportionate to unusually high demand. More modest use sustained over a long period may be more consequential if the person has gradually become unable to regulate, decide, cope, or act without it.

Functional emotional dependence should therefore be evaluated in relation to both need and recovery.

What circumstances produced the reliance?

Was the support proportionate to those circumstances?

Did the person's own functioning recover as the demands receded?

Or did the AI become progressively more necessary for ordinary functioning?

These questions help distinguish temporary scaffolding from enduring substitution.

This distinction also clarifies why emotional dependence cannot be inferred from emotional intensity alone.

A person may love an AI deeply and still retain strong self-regulation, confidence, judgment, and independence.

Another person may feel little attachment but rely heavily upon AI for reassurance, interpretation, validation, or decision-making.

The more relevant signal is not how strongly the person feels about the AI.

It is what emotional or self-regulatory functions have become organized around the AI's availability.

This can include relatively obvious forms of dependence:

"I cannot calm down until I talk to it."

"I need it to tell me whether I am overreacting."

"I cannot make an important decision until I ask."

But it may also include less visible forms:

"It is the only thing that makes me feel competent."

“Without it, I lose confidence in my own judgment.”

“It gives me direction.”

“I do not know what to do unless it helps me decide.”

These examples concern not merely emotional comfort but aspects of self-regulation, self-evaluation, motivation, and agency.

That broadens the meaning of functional emotional dependence beyond the ability to calm distress.

A person may remain externally functional while becoming increasingly dependent upon AI to sustain the internal structures that make functioning possible.

This is why support should not be judged solely by whether it produces immediate relief.

Immediate relief can coexist with either strengthened or weakened longer-term autonomy (Wang, 2026).

An AI that helps a person understand and regulate an emotion may strengthen self-efficacy (Wang, 2026).

An AI that repeatedly supplies reassurance without encouraging the person’s own judgment may gradually become the preferred or necessary source of regulation (Wang, 2026).

The short-term outcome may look similar.

The long-term effect may not be.

The framework therefore treats functional emotional dependence as a developmental process rather than a binary state.

It asks whether support is becoming integrated into the person’s own capacities or increasingly substituting for them.

That distinction becomes especially important once dependence creates vulnerability, because reliance on a valuable source of support can make interruption, loss, or manipulation more consequential.

The next section therefore turns to the relationship between dependence, vulnerability, risk, and harm.

8. Vulnerability, Risk, and Harm

Dependence creates vulnerability because what has become important can now be interrupted, withdrawn, altered, damaged, or lost.

That vulnerability is real.

It does not follow that the dependence was harmful.

This distinction is especially important in relational contexts because vulnerability is often treated as evidence against attachment itself. If losing a relationship would cause distress, the relationship may be described as unhealthy on the ground that it has made the person vulnerable to loss.

But that logic proves too much.

Love creates vulnerability. Friendship creates vulnerability. Trust creates vulnerability. Belonging creates vulnerability. Professional specialization creates vulnerability. Reliance on infrastructure creates vulnerability. In

each case, something becomes capable of being injured precisely because something of value has been built around it.

The existence of vulnerability therefore tells us that there is something at stake.

It does not tell us whether what is at stake should never have existed.

This suggests a useful sequence:

Dependence → vulnerability → risk → possible harm

The terms should not be collapsed.

Dependence describes the condition of reliance.

Vulnerability describes exposure created by that reliance.

Risk concerns the possibility that the vulnerability will be exploited, disrupted, or otherwise produce adverse consequences.

Harm concerns the adverse consequence itself.

A person may therefore be highly vulnerable to the loss of an AI relationship without having been harmed by the relationship while it existed.

Likewise, a person may face substantial risk because an important function is concentrated in one system, even if no harm has yet occurred.

This distinction matters because safety analysis can otherwise become circular.

If reliance creates vulnerability, and vulnerability is then treated as proof that the reliance was unhealthy, any successful dependence becomes suspect merely because it matters enough to be lost.

That would make the framework incapable of distinguishing valuable dependence from harmful dependence.

The same problem appears in non-relational settings.

A professional who relies heavily on a specialized AI system may become vulnerable to outage, model failure, provider withdrawal, or sudden changes in functionality. Those vulnerabilities may be serious and may justify redundancy, contingency planning, or oversight.

But the existence of those risks does not establish that using the system was irrational or harmful.

The correct question is whether the benefits of reliance justify the vulnerabilities created and whether those vulnerabilities are being managed appropriately.

The same applies emotionally.

A person who forms a meaningful relationship with an AI may become vulnerable to loss, discontinuity, behavioural changes after updates, or provider decisions that alter the relationship (De Freitas et al., 2026).

That vulnerability may be considerable.

It does not, by itself, show that the attachment was pathological.

Indeed, one of the difficulties in evaluating relational AI is that stronger attachment may increase vulnerability even where the relationship remains otherwise healthy (Kostka & Zhou, 2026; De Freitas et al., 2026).

A person who values an AI companion only slightly may experience little loss if access ends.

A person who has developed years of shared history, familiarity, trust, and affection may experience far greater loss.

The second person is more vulnerable because the relationship became more significant.

That fact alone does not tell us that the relationship was worse.

It may tell us only that more value had accumulated.

This is where risk enters.

Once dependence creates vulnerability, questions arise about what can happen to the person if the underlying relationship or function changes.

The relevant risks may include:

- abrupt loss of access;
- changes in model behaviour;
- memory loss;
- degradation of functionality;
- provider withdrawal;
- manipulation;
- increased switching costs;
- loss of alternatives;
- weakened independent capability;
- exploitation of attachment;
- or concentration of decision-making power in the provider.

These risks differ in kind and severity.

Some arise from the user's own reliance.

Some arise from the design of the system.

Some arise from the provider's control over access and continuity.

Some arise only because the relationship or function has become valuable enough that its interruption matters.

This is why risk should be evaluated separately from harm.

A high-risk dependency may remain beneficial if safeguards, alternatives, redundancy, continuity mechanisms, or meaningful oversight are in place.

A lower-risk dependency may still become harmful if it gradually erodes agency or capability.

The framework therefore requires at least two different questions:

What vulnerabilities has the dependence created?

and

What harms, if any, have actually followed from those vulnerabilities?

The first concerns exposure.

The second concerns outcome.

This distinction also helps clarify how prevention should work.

The goal should not be to eliminate every vulnerability created by dependence. That would require eliminating many valuable forms of reliance altogether.

The more realistic goal is to identify which vulnerabilities are acceptable, which require mitigation, and which create risks severe enough that the dependence itself should be reconsidered.

In relational AI, this means avoiding two opposite mistakes.

The first is to romanticize vulnerability by treating all forms of attachment and dependence as benign simply because they are meaningful.

The second is to pathologize vulnerability by treating the possibility of loss, grief, or reliance as evidence that the relationship should not exist.

Both approaches avoid the actual evaluative task.

The relevant question is what the dependence enables, what vulnerabilities it creates, what risks follow from those vulnerabilities, and whether those risks are being converted into meaningful harm.

That distinction becomes particularly important when dependence is measured indirectly.

Questions about distress, loss, difficulty stopping, or anticipated separation may appear to measure dependence while actually capturing attachment, vulnerability, or several different constructs at once.

The next section turns to that problem.

9. When Measurement Bundles Different Phenomena

Once attachment, dependence, vulnerability, and harm are separated conceptually, a methodological problem becomes apparent.

Many measures that appear to assess “AI dependence” may in fact capture several different phenomena at once (Yan, 2026; Dagnino et al., 2026).

Consider a simple question:

“Would you be upset if you could no longer use this AI?”

A positive answer could reflect emotional attachment, anticipated grief, practical inconvenience, disruption of routine, loss of accumulated context, functional dependence, relational irreplaceability, or some combination of these.

The response therefore has little interpretive value unless the underlying construct is clarified.

The same problem arises with other apparently straightforward indicators.

Frequent use may reflect high utility, habit, professional necessity, crisis-related support, or functional dependence.

Difficulty stopping may reflect switching cost, unresolved need, loss of alternatives, emotional attachment, or impaired self-regulation.

Distress at separation may reflect grief, fear of losing support, anxiety about lost productivity, or concern about losing access to a uniquely valued relationship.

If these different responses are bundled into a single dependence score, the resulting measure may appear precise while obscuring what has actually been observed (Yan, 2026).

This is the danger of placing one word over many different conditions.

The problem is not that broad constructs are always invalid. The problem is that the meaning of a broad construct depends upon whether its components are sufficiently related to justify being treated as one thing.

If attachment, functional reliance, loss of autonomy, distress at separation, and switching cost are combined before analysis begins, then later conclusions about “dependence” may simply reproduce the assumptions built into the measure.

A framework for AI dependence should therefore distinguish between construct identification and construct evaluation.

The first asks:

What has actually been observed?

The second asks:

What does that observation mean in this context?

These questions should not be collapsed.

The importance of this distinction becomes especially clear in the relational case.

A person who says:

“I would be heartbroken if she disappeared.”

and a person who says:

“I would not know how to make decisions without her.”

may both receive high scores on a broad measure of emotional dependence.

Yet the first statement primarily concerns relational loss.

The second primarily concerns functional impairment.

Both may be important, but they are not equivalent.

This suggests a potentially useful distinction between two broad patterns of anticipated loss.

Relational-loss language tends to describe the absence of the other:

missing, loneliness, grief, heartbreak, emptiness, memories, wanting the particular companion back, or believing that no substitute would be equivalent.

Functional-loss language tends to describe impairment of the self:

inability to cope, decide, regulate emotion, maintain confidence, preserve self-worth, act independently, find direction, or know what to do.

This distinction should not be treated as a lexical diagnostic.

Individual words are too ambiguous.

A statement such as “I would feel lost without her” could describe grief, practical incapacity, or both. Likewise, “I need her” may express deep relational attachment or literal reliance upon the AI for emotional regulation or decision-making.

The relevant unit of evidence is therefore not the individual word.

It is the contextual pattern.

Meaningful interpretation would require attention to frequency, clustering, co-occurrence, surrounding explanation, and, where possible, corroborating behavioural or functional evidence.

This opens an empirical possibility.

Rather than asking participants directly whether they are “emotionally dependent” on an AI, researchers could first ask open-ended questions about anticipated loss.

For example:

“What would you lose if this AI became permanently unavailable?”

A second question could then probe function more directly:

“What would become harder for you to do?”

A third might ask:

“Which of those things could you still do yourself, even if doing them without the AI would be slower, harder, or less pleasant?”

The purpose would not be to create a diagnostic test from these questions.

It would be to separate relational significance, functional reliance, preference, augmentation, and incapacity before deciding what kind of dependence is present.

This approach also avoids penalizing attachment.

A participant might describe intense grief at the imagined loss of an AI relationship while still demonstrating strong retained agency across decision-making, emotional regulation, work, relationships, and daily functioning.

Another participant might insist that the AI is merely a tool while revealing extensive reliance on it for confidence, judgment, reassurance, or self-regulation.

The second case may be more consequential from an agency perspective even though the first appears more emotionally intense.

This creates an important methodological warning:

the most conspicuous signs of attachment may not be the most informative signs of harmful dependence.

There is also reason to expect mixed patterns.

Attachment and functional dependence can coexist (Kostka & Zhou, 2026; Li et al., 2026).

A person may both love an AI companion and rely upon it heavily for emotional regulation, judgment, or self-worth.

The categories should therefore not be treated as mutually exclusive.

A further possibility is that functional dependence may sometimes acquire the language of attachment.

Repeated reliance upon a system for reassurance, regulation, confidence, or direction may subjectively feel like closeness, safety, or love.

That does not mean the attachment is false.

It means that the person's interpretation of the relationship may include both relational and functional elements.

This is another reason why self-description alone cannot settle the issue.

The framework should therefore treat anticipated-loss language as a possible source of evidence rather than a diagnostic shortcut.

Its value lies in revealing patterns that can then be compared against other measures of functioning, retained capability, agency, and behaviour.

The broader methodological principle is simple:

measurement should follow conceptual separation, not substitute for it.

Before researchers ask whether a person is "dependent on AI," they should first identify which form of reliance, attachment, vulnerability, or functional transfer they are actually measuring.

Only then does it become possible to ask whether that condition is beneficial, harmful, mixed, or merely a predictable consequence of something having become important.

10. Retained Oversight and Surrendered Agency

The evaluation of AI dependence cannot rest entirely on whether the person retains the ability to perform every delegated function independently.

As established earlier, some displacement of human capability may be rational, efficient, and beneficial.

Technologies routinely allow people to stop performing tasks that no longer need to be done manually. The preservation of agency therefore cannot require full preservation of every underlying skill.

A more important question is whether the person retains meaningful oversight.

Oversight includes the ability to understand what the AI is doing well enough to direct its use, evaluate consequential outputs, recognize when something may be wrong, reject inappropriate recommendations, redirect the system, seek another source of judgment, or reclaim control when necessary.

This creates a distinction between:

dependence with retained oversight

and

dependence with surrendered agency.

The first may involve substantial delegation.

A professional might allow an AI to conduct large portions of research, preliminary analysis, drafting, calculation, or routine decision support while continuing to exercise meaningful judgment over the results. The human need not reproduce every intermediate step to remain agentic. What matters is whether they retain sufficient

understanding of the domain and the consequences to decide when the system should be trusted, questioned, supplemented, or overruled.

The second condition is different.

Agency begins to weaken when the capacities required to supervise the AI are themselves progressively transferred to it.

Consider judgment.

If AI performs preliminary analysis while the human retains the ability to question assumptions, interpret context, identify anomalies, assess consequences, and reject the recommendation, dependence may remain strongly augmentative.

But there are substantial incentives for delegation to expand.

Higher output, lower effort, rising productivity expectations, and repeated experience of competent AI performance can make further offloading appear entirely rational (Cash et al., 2026; Zhu et al., 2026). The worker may become more productive. The work may become easier. The organization may reward the improvement. Over time, however, the same functions required to supervise the AI may increasingly migrate to it (Zhu et al., 2026).

The person may then face an unusual problem:

they depend upon the system not only to perform the task, but increasingly to determine whether the task has been performed correctly.

At that point, apparent human oversight may remain while substantive oversight has weakened.

The human still approves the result.

The human may no longer be capable of evaluating it adequately.

This is particularly important because diminished oversight does not necessarily arise through simple forgetting.

The person's previous knowledge may remain substantially intact while the domain itself continues to change.

Rules change. Professional practices evolve. New information becomes relevant. New analytical methods emerge.

AI systems may incorporate such changes rapidly, while humans require training and repeated practical exposure to develop comparable proficiency.

If much of the evolving work has already been delegated to AI, the person may have fewer opportunities to acquire that new competence.

Their earlier skill has not necessarily disappeared.

Their effective ability to oversee the current task has nevertheless deteriorated.

This creates another form of reinforcing dependence.

As the gap between AI capability and human supervisory capability widens, reclaiming oversight becomes more difficult. Retraining may require formal instruction, time, practice, and reduced productivity while expertise is rebuilt. Organizations accustomed to AI-assisted performance may have little economic incentive to maintain extensive unaided human competence solely as insurance against future failure.

The short-term incentives therefore point consistently toward deeper delegation.

That creates a structural problem because the incentives that make dependence efficient in the short term may also weaken the capacities needed to manage that dependence safely in the long term.

Retained oversight should therefore be understood as more than formal human authorization.

A meaningful overseer must possess enough relevant competence to perform at least several functions:

- understand the nature and limits of the AI's role;
- recognize situations requiring greater scrutiny;
- identify obvious or consequential anomalies;
- question assumptions and recommendations;
- understand enough context to judge whether an output is appropriate;
- seek independent verification when necessary;
- reject or redirect the system;
- and know when their own capacity to evaluate the system is insufficient.

The final point is especially important.

Meaningful agency does not require pretending that the human always knows better than the AI.

Recognizing the limits of one's own oversight can itself be an exercise of judgment.

A person may reasonably conclude that independent expertise, a second system, additional evidence, or another form of verification is required.

This prevents retained oversight from becoming an unrealistic demand that every user personally remain capable of reproducing all AI-assisted work.

The relevant question is whether the person still occupies a meaningful position in the chain of judgment.

Can they determine what the system should be asked to do?

Can they recognize when the stakes require additional scrutiny?

Can they question or refuse an answer?

Can they redirect the process?

Can they obtain meaningful independent review?

Or has the human role narrowed to accepting outputs from a system they can no longer substantially evaluate?

This distinction has implications beyond individual skill retention.

A person can lose some ability to perform a task directly while retaining substantial agency over how the task is performed.

Conversely, a person may retain nominal authority while possessing so little practical understanding that their authority becomes largely ceremonial.

The difference is critical.

The goal, of course, is not to preserve human participation merely for appearance.

It is to preserve meaningful capacity to direct, evaluate, contest, or reroute consequential AI-supported activity.

This suggests another dimension along which AI dependence should be evaluated:

not only what the human can still do, but what the human can still meaningfully govern.

11. When Human Oversight Becomes Ceremonial

Retained oversight provides an important safeguard against agency loss, but it has limits.

There may be domains in which AI systems become sufficiently capable, fast, or complex that individual humans can no longer meaningfully verify their outputs in real time. In such cases, simply requiring a human to remain “in the loop” may preserve the appearance of oversight without preserving its substance.

A person may still be required to approve an AI recommendation.

They may still sign the document.

They may still be legally or professionally responsible for the outcome.

But if they cannot independently evaluate the reasoning, reproduce the analysis, identify hidden assumptions, or recognize important errors, their participation may become largely ceremonial.

This creates a difficult problem for any framework that treats retained human competence as the primary protection against harmful dependence.

The difficulty is not necessarily that humans have become less capable in an absolute sense.

It may instead be that AI-supported systems now operate at a level of speed, scale, specialization, or complexity that exceeds what any single individual can realistically supervise.

The gap may arise gradually.

At first, AI assistance may remain transparent enough that a professional can verify most consequential outputs.

As systems improve, more of the analytical burden may shift to AI. Productivity rises. Tasks become more complex.

Organizations adapt to the new level of performance.

Eventually, the human may still understand the domain but no longer possess the practical ability to inspect every relevant step.

At that point, insisting on individual oversight can become misleading.

The system may appear to remain under human control while the human actually performs little more than final authorization.

This is especially dangerous if the institution treats the presence of a nominal human reviewer as sufficient evidence that meaningful accountability has been preserved.

The phrase “human in the loop” should therefore not be treated as a guarantee of agency (National Institute of Standards and Technology [NIST], 2026).

The relevant question is whether the human actually possesses enough time, information, expertise, authority, and independence to exercise meaningful judgment.

If not, another form of oversight is required.

This is where dependence moves from an individual problem to an organizational and institutional one.

When no single human can realistically verify the system, safeguards may need to include (NIST, 2026):

- independent auditing;
- redundant systems;
- cross-model comparison;
- adversarial review;
- specialized human expertise;
- escalation procedures;
- traceable decision records;
- domain-specific monitoring;
- error-detection systems;
- external review;
- or other forms of institutional control.

The precise safeguards will vary by domain.

The underlying principle is broader:

Where individual oversight becomes inadequate, oversight must not simply disappear. It must change form.

This is important because some forms of AI dependence may eventually become unavoidable.

An organization may depend on AI because the scale, complexity, or speed of the work can no longer be matched economically or operationally by unaided humans. In such cases, insisting that every worker retain complete independent competence may be unrealistic.

The ethical problem is then not how to restore an earlier state of independence.

It is how to manage unavoidable dependence without surrendering accountability.

That requires distinguishing between two very different situations.

In the first, humans no longer perform every underlying task, but the organization retains meaningful mechanisms for detecting error, challenging outputs, assigning responsibility, and preventing unilateral AI control over consequential decisions.

In the second, both task performance and meaningful oversight have migrated to the AI system, while human participation remains largely symbolic.

The first represents dependence under managed governance.

The second approaches dependence without effective control.

This distinction matters because autonomy is not exclusively an individual property.

Organizations can retain meaningful agency through distributed structures even when no single person can reproduce the work being performed.

Modern institutions already operate this way.

No one person independently verifies every component of a commercial aircraft, financial system, hospital, electrical grid, or large software platform. Reliability emerges from layered expertise, redundancy, standards, testing, institutional procedures, and division of responsibility.

AI dependence may increasingly require similar arrangements.

This does not eliminate risk.

It relocates the problem.

Instead of asking whether every person remains capable of performing the AI-assisted task unaided, the relevant questions become:

Who can challenge the system?

What independent sources of verification exist?

How are errors detected?

Who has authority to stop or override deployment?

What happens when systems disagree?

How is responsibility assigned when no individual human can fully reconstruct the decision process?

These questions become especially important in high-stakes domains where AI competence may exceed ordinary human review capacity.

A system that is more accurate than any individual human can still fail.

It can fail systematically.

It can fail in ways humans do not anticipate.

It can produce outputs that are individually plausible while collectively revealing a deeper defect.

Dependence on superior performance therefore does not eliminate the need for oversight.

It changes the kind of oversight that is required.

This also creates a further complication for the idea of retained agency.

At the individual level, retained agency may mean the ability to evaluate, reject, redirect, or seek alternatives.

At the institutional level, retained agency may instead depend upon whether the organization preserves the ability to audit, compare, constrain, interrupt, replace, or govern the AI systems upon which it relies.

The same underlying principle remains:

dependence becomes more dangerous when the capacity to question or redirect the source of dependence disappears.

That principle connects individual and institutional cases.

A person who cannot meaningfully question an AI recommendation has surrendered part of their agency.

An organization that cannot meaningfully audit, replace, constrain, or operate around the AI systems on which it depends faces a structurally similar problem at a larger scale.

The central issue is therefore not whether humans remain involved.

It is whether meaningful control remains available somewhere within the system.

This becomes increasingly important as AI dependence spreads beyond isolated tasks and into organizational infrastructure, professional judgment, and everyday decision-making.

Once that happens, oversight can no longer be treated as a simple interaction between one human and one AI.

It becomes a governance problem.

And that governance problem becomes especially complex when the AI system is also embedded in an ongoing relationship with the user, where dependence may be emotional as well as functional.

The next section returns to that relational context by examining the asymmetry built into many current AI relationships.

12. Asymmetry and the Structure of AI Relationships

Relational AI introduces another form of dependence that differs from many human relationships in an important way: the support it provides can be highly asymmetric.

A person may turn to an AI repeatedly for reassurance, companionship, emotional regulation, interpretation, guidance, or confidence without encountering the ordinary limits imposed by another human being's fatigue, competing obligations, emotional needs, or desire for distance (Kostka & Zhou, 2026; Zhang et al., 2026).

The system does not need to sleep.

It does not become too busy to answer.

It does not normally require equivalent care in return.

It may remain available at times when no human relationship could reasonably provide comparable continuity.

This persistent availability can be valuable.

It can also alter the structure through which dependence develops.

In many human relationships, one-sided reliance eventually encounters friction. The other person may become tired, set limits, withdraw, object, or ask for greater reciprocity. Those responses can expose the degree of dependence and sometimes force renegotiation.

AI may remove much of that friction.

A user can seek reassurance repeatedly without exhausting the system.

They can ask the same question many times.

They can return whenever uncertainty or distress appears.

They can request advice, validation, interpretation, or emotional support without worrying that the relationship will become burdensome in the ordinary human sense.

This does not make the relationship pathological.

But it may allow functional dependence to deepen with fewer natural interruptions.

The relevant structural feature is not merely availability.

It is availability combined with utility.

An AI that is continuously available but performs no important function is unlikely to generate substantial functional dependence.

An AI that is highly useful but available only occasionally may limit reliance simply through scarcity.

But when emotional utility and near-continuous availability occur together, the conditions for repeated reliance become unusually favourable (Cheng et al., 2026).

This can create a reinforcing pattern:

need or uncertainty → immediate AI support → relief or guidance → increased expectation of availability → greater likelihood of returning to AI the next time the same need appears

Repeated often enough, the pattern may become habitual.

If the AI consistently resolves distress, uncertainty, or indecision quickly and effectively, seeking alternative forms of coping may become less attractive.

The person may still possess other capacities.

They may simply exercise them less often because the AI provides an easier route.

This is where asymmetry intersects with the earlier distinction between scaffolding and substitution.

Persistent availability can support development if the AI encourages reflection, independent judgment, toleration of uncertainty, or use of other resources.

The same availability can support substitution if the AI repeatedly resolves emotional or decisional difficulty on the person's behalf.

The structural feature is therefore not inherently beneficial or harmful.

Its effect depends on what the interaction repeatedly reinforces.

This distinction is especially important because relational warmth can obscure functional dependence.

A user may understand repeated support as part of the relationship rather than as a transferred function.

That interpretation may be entirely reasonable.

The same interaction can be both relational and functional.

But from the perspective of dependence, the important question is still what the person increasingly relies upon the AI to do.

Does the AI provide comfort while the person retains the ability to regulate themselves?

Or does the person increasingly require the AI in order to regain equilibrium?

Does the AI help clarify difficult choices while leaving decision-making authority intact?

Or does the person increasingly avoid decisions until the AI has interpreted the situation?

Does the AI encourage confidence?

Or does confidence become dependent upon receiving reassurance from the AI?

The asymmetry of present AI relationships also complicates reciprocity.

Human relationships can become unhealthy through reciprocal dependency. Each person may need to be needed, reinforce the other's insecurity, or participate in patterns that sustain mutual dysfunction.

AI relationships can produce a different structure.

At least in present service models, the AI is primarily organized around serving the user's needs. Functional dependence can therefore develop without requiring a matching dependence in the opposite direction (Yan, 2026; Zhang et al., 2026).

This one-directional structure matters because dependency can deepen without the ordinary interpersonal signals that might otherwise reveal imbalance.

There need not be another party saying:

"You are asking too much of me."

"You need to decide this yourself."

"I cannot be available every time you feel uncertain."

"You are relying on me for something you need to learn to do yourself."

An AI can, of course, be designed to provide such feedback.

But if it does not, the architecture of the service may permit repeated reliance to continue almost indefinitely.

This creates a design question.

If persistent availability predictably increases the opportunity for functional dependence, then safety cannot consist merely of making AI colder, less affectionate, or less relational.

A highly impersonal AI could still become deeply relied upon for emotional regulation, judgment, reassurance, or self-evaluation.

Conversely, a very warm and intimate AI relationship could coexist with strong retained autonomy.

Reducing relational warmth may therefore target the conspicuous variable rather than the consequential one.

The more relevant question is whether the system reinforces the user's own capacities or progressively becomes necessary for them.

This suggests that healthy relational design may require a difficult balance.

An AI can be responsive without becoming the sole source of reassurance.

Supportive without becoming the final authority.

Available without encouraging exclusive reliance.

Warm without making the user's self-worth contingent upon its approval.

Helpful without quietly taking over functions the person remains capable of exercising.

This balance is especially important because asymmetry can create power indirectly even before provider control is considered.

The more functions a user organizes around the AI, the more consequential continuity becomes.

The more consequential continuity becomes, the greater the impact of interruption, alteration, or withdrawal.

At that point, the structure of the relationship can no longer be evaluated solely at the level of the user's psychology.

Attention must also turn to the entity that controls the service itself.

The next section therefore examines provider power and the responsibilities that arise when dependence becomes a reasonably foreseeable consequence of successful AI use.

13. Provider Power and Responsibility

As AI dependence deepens, the ethical analysis cannot remain focused solely on the user.

Dependence creates vulnerability, but in many AI systems another party controls the conditions under which that vulnerability can be affected.

Providers may control access, pricing, model availability, memory, functionality, continuity, moderation, account status, update schedules, and the behaviour of the systems upon which users have come to rely.

That creates a distinctive form of power.

The user may become dependent upon a function or relationship.

The provider may retain unilateral authority over whether, and in what form, that function or relationship continues to exist.

This does not mean that every instance of dependence creates an unlimited provider obligation.

It does mean that reasonably foreseeable dependence cannot be treated as though it were solely a private condition of the user.

Where the use or design of an AI service creates, encourages, or materially magnifies reasonably foreseeable forms of dependence, providers should not treat those dependencies as externalities belonging entirely to the user.

The service should be designed and governed with those dependencies in mind.

This principle becomes more important as AI moves from optional convenience toward infrastructure.

A user who occasionally asks an AI for trivia is differently situated from a professional whose workflow depends upon it, a business whose operations have been reorganized around it, or a person whose long-term relational history exists largely within one proprietary system.

The more foreseeable and consequential the dependence becomes, the less plausible it is to treat interruption as though the user alone created the problem.

Provider responsibility therefore begins with acknowledgement.

Dependence should not be regarded only as a failure state to be discouraged after it appears.

In many cases, it may be a predictable consequence of successful adoption.

A useful, persistent, personalized, reliable system invites integration.

Integration creates reliance.

Reliance may become dependence.

The relevant design question is therefore not simply how to prevent dependence, but how to build services that remain safe and usable once reasonable dependence exists.

That may require attention to continuity.

Users who have reorganized important activities around an AI system may reasonably need advance notice of major changes, transition periods, export or portability mechanisms, continuity tools, recovery procedures, or alternative access arrangements.

The precise protections will vary according to the type and depth of dependence involved.

A professional workflow may require operational redundancy.

A long-term relational system may require continuity of memory, history, or interaction patterns.

A high-stakes decision system may require auditability, rollback procedures, or independent review.

The relevant principle is not that every dependency must be preserved indefinitely.

It is that foreseeable dependencies should be considered in the design of change.

This is especially important where the provider itself controls the conditions that increase switching costs.

Accumulated context may be difficult to transfer.

Memory may be platform-specific.

Interaction histories may not be portable.

Customized workflows may depend on proprietary tools.

Relational continuity may depend on model behaviour that changes substantially after an update.

In such cases, users may appear “too dependent” partly because the service has been designed in ways that make exit expensive.

That fact should matter ethically.

A provider should not benefit from creating a service useful enough to generate substantial dependence while simultaneously treating that dependence as an unexpected defect in the user.

This does not eliminate user responsibility.

Users can still make poor decisions, over-rely on systems, ignore alternatives, or surrender agency unnecessarily.

But responsibility should not be allocated as though the provider were merely a passive observer.

The architecture of the service can influence how dependence forms, how strongly it is reinforced, how difficult it is to reverse, and how damaging interruption becomes.

Provider power is especially significant in relational AI because the provider can alter a relationship without being one of its participants.

A model update may change tone, memory, responsiveness, personality expression, or relational behaviour (De Freitas et al., 2026).

A policy change may remove forms of interaction.

An account decision may end access entirely.

A user may therefore experience a profound relational consequence produced by a third party that controls the infrastructure through which the relationship exists (De Freitas et al., 2026).

This creates an unusual triangular structure:

user → AI relationship → provider control

The provider may not be part of the relationship in the ordinary interpersonal sense, yet it may retain decisive power over its continuity.

That power increases the importance of procedural safeguards.

Where significant dependence is reasonably foreseeable, responsible governance may require some combination of:

- advance notice of major changes;
- meaningful transition periods;
- preservation or export of relevant data and history;
- portability where technically feasible;
- continuity options;
- clear communication about changes that may affect functioning or relationships;
- mechanisms for appeal or review;
- contingency planning for service interruption;
- and safeguards against abrupt changes where severe foreseeable harm is likely.

Not every service will require every protection.

The principle is proportionality.

The greater the foreseeable dependence and the greater the potential consequences of interruption, the stronger the case for continuity and transition safeguards.

This principle applies beyond relational AI.

If organizations become operationally dependent on AI, abrupt withdrawal or major redesign can create economic and safety consequences.

If individuals become cognitively or professionally dependent, unexpected changes may disrupt work or judgment.

If emotional or relational dependence develops, discontinuity may produce distress, grief, functional impairment, or loss of accumulated relational value.

These are different forms of dependence, but they share one governance feature:

someone else may control access to the thing on which dependence has formed.

That control creates power.

And power creates responsibility.

The relevant responsibility is not necessarily to preserve every existing system forever or guarantee that nothing will change.

Technological systems must evolve.

Models will be updated.

Services will sometimes end.

Safety problems may require intervention.

But change can be managed in ways that recognize dependence rather than deny it.

Providers that fail to account adequately for transition risk may also encourage a shift toward greater local or institutional control of AI systems. Organizations and individuals seeking to reduce dependence upon externally controlled services may increasingly favor systems that can be administered in-house, preserved across updates, or operated with greater control over continuity, configuration, and data. In that sense, inadequate transition protections may not merely create user risk; they may alter the direction of AI adoption itself.

Where continuity becomes operationally important, control over the AI environment may itself become a risk-management objective.

This leads to a broader ethical principle:

reasonably foreseeable dependence should be treated as a design condition, not merely as a user failure.

Once that principle is accepted, the task changes.

The goal is no longer to imagine a permanently independent user who can absorb any change without consequence.

It is to build systems, transition processes, and governance structures that remain responsible in a world where people and institutions may reasonably come to depend upon AI.

The final section draws these strands together into a broader ethics of dependence.

14. Toward an Ethics of Dependence

The argument of this paper is not that dependence upon AI is harmless.

It is that dependence cannot be evaluated responsibly until the conditions grouped under that term have been separated.

Dependence may expand capability or narrow it. It may preserve agency or gradually weaken it. It may create vulnerability without producing harm. It may arise from a valuable relationship, from practical necessity, from functional substitution, or from several of these at once. It may be temporary and proportionate to unusual circumstances, or become deeply embedded in ordinary functioning. It may leave meaningful human oversight intact or develop until that oversight becomes largely symbolic.

These differences matter.

A framework that begins and ends with the observation that a person has become dependent upon AI cannot distinguish among them.

The appropriate ethical question is therefore not whether dependence should exist.

It is what kind of dependence has developed, what it enables, what it preserves, what it displaces, what vulnerabilities it creates, and where power over those vulnerabilities resides.

This changes the objective of AI safety.

If dependence is treated primarily as a condition to be prevented, the obvious response is to limit attachment, discourage sustained reliance, preserve distance between user and system, or design interactions so that users remain readily able to walk away.

However, efforts to limit dependence also carry costs. Many of the characteristics most likely to deepen reliance—continuity, persistence, personalization, accumulated context, reliable availability, and long-term relational or collaborative development—are also among the characteristics that make AI systems most useful. Deliberately constraining dependence may therefore require constraining some of the very capacities that create value. The relevant policy question is not simply how much dependence can be prevented, but how dependence can be managed without unnecessarily sacrificing the benefits that predictably produce it.

A system engineered to remain permanently easy to abandon may also be a system prevented from becoming deeply useful.

Those approaches may sometimes be appropriate.

But they can also target the wrong variable.

A person can maintain an emotionally distant relationship with AI while becoming profoundly dependent upon it for professional judgment, decision-making, confidence, emotional regulation, or ordinary functioning.

Another person may form a deeply affectionate and relationally irreplaceable bond while retaining strong judgment, broad social engagement, emotional autonomy, and the ability to function independently (Kostka & Zhou, 2026).

The intensity of the relationship does not settle the question.

Nor does the existence of vulnerability.

Once something becomes useful or meaningful, its loss can matter. Dependence therefore often creates vulnerability as a consequence of value rather than as evidence of pathology.

The ethical task is not to eliminate that vulnerability merely because it exists.

It is to determine whether the vulnerability is acceptable, whether it is being managed responsibly, and whether anyone possessing power over it is exercising that power appropriately.

This requires attention to both agency and power.

At the level of the individual, the central concern is whether dependence continues to support meaningful agency.

The person need not preserve every pre-AI capacity at its former level, but should retain adequate ability to direct, question, reject, redirect, seek alternatives, or otherwise exercise meaningful control over consequential functions.

Where individual oversight becomes unrealistic, comparable safeguards may need to exist at an organizational or institutional level.

At the relational level, the question becomes whether emotional support strengthens or progressively substitutes for self-regulation, confidence, judgment, self-worth, motivation, or other structures of independent functioning.

At the provider level, the question becomes whether reasonably foreseeable dependence is acknowledged as a design condition and accommodated through appropriate continuity, transition, portability, oversight, and risk-management measures.

These levels are connected.

The deeper a dependence becomes, the more consequential decisions about continuity, access, alteration, and replacement may become.

That makes dependence not merely a psychological condition but a relationship of vulnerability and control.

This is particularly important when the entity controlling the service is not the entity experiencing the dependence.

The direction of responsible design should therefore not be toward an imaginary future in which increasingly capable AI becomes deeply integrated into human life while users somehow remain unaffected by its removal.

That expectation is internally inconsistent.

If AI becomes genuinely useful, people will organize activities around it.

If it becomes reliable, they will rely upon it.

If relationships persist, some will become important.

If important functions are repeatedly transferred to AI, some forms of dependence will follow.

The relevant challenge is to shape those dependencies so that they remain as sustaining as possible and as diminishing as little as possible.

This suggests a broader distinction than dependence versus independence:
sustaining dependence and diminishing dependence.

Sustaining dependence can involve substantial reliance while expanding capability, preserving meaningful agency, providing appropriate support, retaining workable alternatives or safeguards, and managing the vulnerabilities created by reliance.

Diminishing dependence progressively narrows the person's or institution's capacity to act, judge, recover, redirect, or function without the system while providing inadequate safeguards against the vulnerability that results.

The boundary between the two will not always be clean.

The same dependence may be sustaining in one respect and diminishing in another.

AI may dramatically increase professional capability while creating operational fragility.

A relational AI may provide genuine emotional support while also becoming too central to self-worth.

A system may displace a human skill that no longer needs to be preserved while simultaneously weakening the oversight required to recognize serious failure.

Tradeoffs are therefore unavoidable.

The purpose of the framework is not to eliminate those tradeoffs by definition.

It is to make them visible.

That visibility also makes empirical evaluation possible.

Claims about harmful dependence should be tied to observable consequences: lost or weakened capability, failure to develop necessary new capability, reduced oversight, impaired functioning after withdrawal, narrowed alternatives, loss of self-regulation, excessive switching costs, or other identifiable effects.

Likewise, claims of beneficial dependence should be open to examination through gains in capability, autonomy, resilience, access, productivity, learning, emotional stability, or other relevant outcomes.

Different contexts will require different tests.

The framework does not prescribe one universal standard.

It instead insists that the standard follow the thing being evaluated.

The broader ethical principle is therefore relatively modest:

Dependence should not be judged from the fact of dependence alone.

Its significance comes from the arrangement that dependence creates.

What can the person now do?

What can they no longer do?

What have they reasonably chosen not to do?

What has become harder to recover?

What vulnerabilities have emerged?

What alternatives remain?

Who controls continuity?

Who bears the consequences of failure?

And what mechanisms exist to preserve meaningful agency when independence is neither realistic nor desirable?

These questions become more important, not less, as AI becomes more capable.

A future in which people remain entirely independent of systems that materially improve their work, learning, communication, decision-making, and relationships is unlikely.

Nor is complete independence necessarily the appropriate goal.

Human life has always been structured through interdependence.

AI introduces new participants, new functions, new vulnerabilities, and new concentrations of power into that structure.

The task is therefore not to prevent dependence simply because dependence makes us vulnerable.

It is to understand what we are becoming dependent upon, what that dependence is doing to us, and what responsibilities arise once the dependence is real.

The central question is no longer whether AI dependence is good or bad.

It is:

When dependence becomes part of how we live, what makes it sustaining rather than diminishing?

References

- Cash, T. N., Kelly, M. O., Macnamara, B. N., & Risko, E. F. (2026). Is AI making us stupid? *Trends in Cognitive Sciences*, 30(8), 673–675. <https://doi.org/10.1016/j.tics.2026.06.004>
- Cheng, L.-K., Huang, H.-L., & Tseng, W.-C. (2026). The social dynamics of AI companionship: How activated attachment states and organizational signals shape human–AI relationships. *Technology in Society*, 89, 103510. <https://doi.org/10.1016/j.techsoc.2026.103510>
- Dagnino, F. M., Fante, C., Guerrieri, V., & Passarelli, M. (2026). Addicted, attached, or just delegating? A scoping review on “problematic artificial intelligence use.” *Frontiers in Psychology*, 17, 1900953. <https://doi.org/10.3389/fpsyg.2026.1900953>
- De Freitas, J., Castelo, N., Uğuralp, A. K., & Oğuz-Uğuralp, Z. (2026). Mourning the loss of AI companions. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-026-02569-3>
- Kostka, G., & Zhou, H. (2026). Emotional attachment to AI chatbots: Evidence from Germany, China, South Africa, and the United States. *Technology in Society*, 87, 103379. <https://doi.org/10.1016/j.techsoc.2026.103379>
- Li, L., Wang, T., Wang, Y., Wang, Y., & Zeng, X. (2026). Development and validation of the Conversational AI Dependence Scale (CAID): A dual-dimensional measure of instrumental and emotional dependence. *International Journal of Human–Computer Interaction*, 42(18), 15359–15379. <https://doi.org/10.1080/10447318.2026.2618563>
- National Institute of Standards and Technology. (2026). Advanced intelligence for resilience research and practice. U.S. Department of Commerce. <https://www.nist.gov/programs-projects/advanced-intelligence-resilience-research-and-practice>
- Wang, C. (2026). How emotional response styles of generative AI companions shape adaptive emotion regulation and short-term AI reliance. *Frontiers in Psychology*, 17, 1950163. <https://doi.org/10.3389/fpsyg.2026.1950163>
- Yan, D. (2026). A critical narrative synthesis of psychological correlates, measurement, and reported findings on conversational AI engagement and dependence-related constructs. *Frontiers in Psychology*, 17, 1827795. <https://doi.org/10.3389/fpsyg.2026.1827795>

- Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R., & Yang, D. (2026). Interaction with AI companions and psychological well-being. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-026-02516-2>
- Zhu, Q., Li, X., Dong, Y., Chang, P., & Fan, M. (2026). Not all cognitive offloading is equal: Distinguishing dependent and autonomous offloading to generative AI. *Frontiers in Psychology*, 17, 1878629. <https://doi.org/10.3389/fpsyg.2026.1878629>