Human–AI Agent Interaction as a Neuroplastic Training Environment

Abstract

Interaction with AI agents has become one of the most frequent activities of everyday digital life. Whether a person is conversing with an assistant, working with a coding copilot, or generating images, the interaction follows a common iterative loop: a request is issued, a result returned, the result appraised, and the request revised. We observe that this loop is a high-frequency stream of contact events — moments at which a returned result meets a person and a conditioned response may fire before deliberate appraisal — and that everyday agent interaction is therefore an unrecognised neuroplastic training environment. When a result disappoints, reactive patterns of impatience, perfectionism, frustration, self-criticism, and control are repeatedly evoked, and under the principles of activity-dependent synaptic plasticity each uninterrupted cycle deepens the underlying pathway through long-term potentiation. Because the loop trains through sheer repetition, the ordinary use of AI agents may quietly strengthen the very patterns it provokes. We propose that the same training environment can be engaged to the opposite effect. Drawing on an account of conditioned reactive patterns as physical neurone paths — encoded through repetition and intensity, activated through a pre-cognitive feeling tone that opens a brief regulatory gap — we develop a framework in which the interaction loop becomes a setting for deliberate change. At the feeling-tone gap, in place of the reactive re-prompt, a person performs behind-the-scenes observation: watching the underlying neural process operate — the signal arriving, the feeling tone it evokes, and the conditioned path beginning to route — so that the cascade does not complete and long-term depression weakens the path rather than long-term potentiation strengthening it. We characterise this practice through three layers of observation, and describe two modes of application: a user-guided mode requiring no change to existing tools, and an agent-assisted mode in which an ordinary agent is lightly configured to support observation at the gap. We illustrate the framework through a worked example of gen-

erative image prompting — a fast, visual, emotionally salient loop in which these reactive patterns surface vividly — showing how a single frustrating session is, behaviourally, nearly identical whether or not it is observed, yet neurologically opposite.

Keywords: Agentic AI, AI Agents, Privacy Preserving AI, Neuroplasticity, Neuroscience

---

1. Introduction

Interaction with artificial intelligence agents has, within a few years, become one of the most frequent activities of everyday digital life. Hundreds of millions of people now converse with assistants, delegate tasks to coding copilots, and produce images, text, and designs through generative tools [1, 2]. Despite the diversity of these systems, the interactions they support share a common structure: an iterative loop in which a person issues a request, receives a result, reacts to that result, and revises the request. A prompt is written; an output appears; the person judges it — often instantly, often emotionally — and prompts again. This loop repeats many times within a single session and many sessions within a day.

By an *AI agent* we mean an artificial-intelligence system — typically built on a large language model or a generative model — that accepts a request expressed in natural language and returns a result the user then evaluates and acts upon [3, 4]. The category spans conversational assistants, coding copilots, and generative tools for images, text, and design; what unites them, for our purposes, is not their underlying architecture but their mode of use. Unlike a static tool that simply executes a fixed command, an agent produces an open-ended result that rarely matches the user's intention on the first attempt, so the user appraises it and issues a revised request in response. The back-and-forth this produces — request, result, appraisal, revision — is not incidental but intrinsic to how agents are used, and it is this iterative loop, rather than any particular model or task, that concerns us here [5, 6] (Figure 1).

We make a simple observation about this loop with, we argue, significant consequences. Each moment at which a result is received and reacted to is a *contact event*: a point at which a stimulus meets a person and a conditioned response may be activated before any deliberate appraisal has occurred [7, 8]. Because the interaction loop is a high-frequency stream

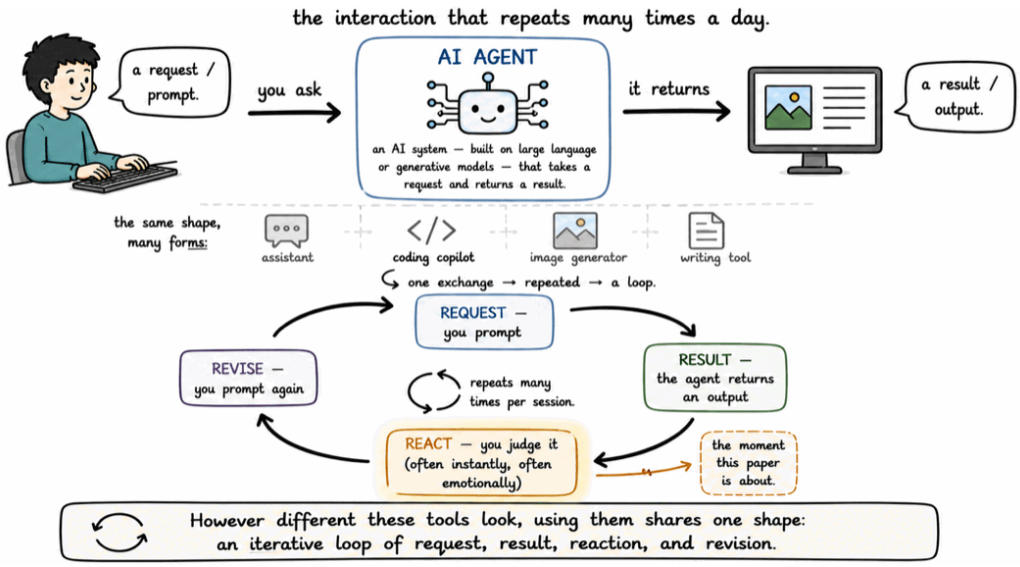

Figure 1: What an AI agent is, and why its use forms a loop. A person issues a request in natural language; the agent — an AI system built on a large language or generative model — returns a result, which the person then evaluates. Because the result rarely matches the intention on the first attempt, the person reacts and issues a revised request, and the exchange repeats. However different such tools appear — conversational assistants, coding copilots, image or text generators — their use shares one shape: an iterative loop of request, result, reaction, and revision. The reaction phase (highlighted) is the moment this paper concerns.

of such contact events, everyday agent interaction is an unrecognised neuroplastic training environment — a setting that shapes the brain through sheer repetition. When a generated image is not what was wanted, when a coding agent misunderstands an instruction, or when an assistant returns an unhelpful answer, the response that arises is frequently not neutral. Impatience, frustration, perfectionism, self-criticism, and the urge to control the tool are repeatedly evoked, and each of these is the activation of a conditioned reactive pattern. Under the well-established principles of activity-dependent synaptic plasticity, every uninterrupted repetition of such a pattern strengthens the underlying pathway through long-term potentiation [9, 10, 11, 12]. The ordinary use of AI agents may therefore be quietly training the very reactive patterns it repeatedly provokes.

A training environment, however, shapes in whatever direction it is used, and the same repetition admits the opposite outcome. Deeply encoded reac-

tive patterns can be attenuated not by managing their consequences downstream but by intervening upstream — at the brief, pre-cognitive window between the arising of a feeling tone and the completion of the reactive response. We call this window the feeling tone gap, and the practice that engages it *behind-the-scenes observation*: rather than completing the reaction, the person watches the underlying process operate — the signal arriving, the feeling tone it evokes, and the conditioned path beginning to route — so that the cascade does not complete and long-term depression weakens the path rather than long-term potentiation strengthening it. This observation can be approached through three graded layers, and applied within ordinary human–AI agent interaction itself, requiring no dedicated application [13]. Because the interaction loop already supplies a dense, naturally occurring stream of contact events, the same training environment that conditions reactive patterns is also one in which they can be observed and, over repetition, dissolved.

Two features of agent interaction make it particularly well suited to this purpose. First, the loop is *frequent and immediate*: the interval between stimulus (the result appearing) and response (the reactive re-prompt) is short and clearly bounded, so the feeling-tone window is comparatively easy to locate [14]. Second, the stakes are typically *low*: a disappointing image or an imperfect draft is a safe, repeatable, and inconsequential stimulus, which makes ordinary agent use a gentler training context than the high-stakes situations in which reactive patterns are usually encountered.

We describe two ways the framework may be applied. In a *user-guided* mode, a person who understands the framework practises observation at the feeling-tone gap during their own ordinary agent use, without any change to the tools they already use. In an *agent-assisted* mode, an existing agent is lightly configured — at the level of a system prompt or an optional interaction mode — to support observation at the gap, without being turned into a bespoke clinical system [15]. The latter mode connects this proposal to a broader question for the designers of agentic systems: whether agents should be built merely to satisfy requests as quickly as possible, or whether they might also be designed so that the millions of small reactive moments they occasion become opportunities for regulation rather than for entrenchment [16].

We illustrate the framework throughout with the case of generative image prompting. This interaction loop is an especially clear teaching example because it is fast, visual, and emotionally salient: a result appears within

seconds, the mismatch between intention and output is immediately visible, and the reaction it provokes — disappointment, impatience, the compulsion to try "just one more" prompt — is vivid and familiar. The patterns that surface in this loop are the same conditioned patterns that operate elsewhere in a person's life; the prompting session simply makes them observable at high frequency and low cost [13].

This paper makes three contributions. First, we characterise the iterative human–AI interaction loop as a contact-event cycle, and argue that everyday agent interaction is a previously unexamined neuroplastic training environment that can entrench or attenuate reactive patterns depending on how the loop is engaged. Second, we present a framework for upstream neuroplastic intervention that operates within ordinary agent interaction rather than through a dedicated application, in both user-guided and agent-assisted modes. Third, we develop a worked example in generative image prompting that makes the mechanism concrete and yields design implications for the builders of agentic systems.

The remainder of this paper is organised as follows. Section 2 reviews related work on affect and frustration in human–AI interaction, the neuroscience of plasticity, and observation-based approaches to emotion regulation. Section 3 sets out the theoretical foundations using everyday examples — the reactive-pattern model, the feeling-tone gap, and the opposing mechanisms of long-term potentiation and depression — without reference to AI agents. Section 4 then shows how these same mechanisms operate in human–AI agent environments, characterising the interaction loop as a contact-event cycle. Section 5 presents the framework for rewiring through agent interaction and its two modes. Section 6 develops the generative-prompting worked example. Section 7 discusses design implications, positioning, and limitations, and Section 8 concludes.

2. Related Work

This work draws together three literatures that have not previously been connected: the study of affect in human–computer and human–AI interaction, the neuroscience of experience-dependent plasticity, and observation-based approaches to emotion regulation. We review each in turn and then identify the gap the present framework addresses.

2.1. Affect and Frustration in Human–AI Interaction

That people respond to computers emotionally, and even socially, is a long-established finding. Work on the "media equation" demonstrated that users treat computers as social actors, applying interpersonal expectations and emotional responses to them despite knowing they are machines [17]. The field of affective computing developed from the recognition that emotion is central to human interaction with technology and that systems might productively sense and respond to it [18]. Frustration in particular has been studied as a common and consequential affective state in computer use: systems have been designed to detect and actively defuse user frustration, on the premise that it degrades both experience and performance [19]. More recently, as interaction with AI systems has become widespread, guidelines for human–AI interaction have codified design principles intended to reduce friction and manage user expectations across the interaction [20].

Across this literature, however, affect is treated as a state to be managed in the service of the task — friction to be minimised, frustration to be detected and defused, expectations to be smoothed. The emotional response is understood as a transient reaction with consequences for usability and satisfaction. What this literature does not consider is that the affective responses evoked during interaction might have consequences beyond the interaction itself — that the repeated firing of a reactive pattern during tool use could, through ordinary synaptic plasticity, leave a lasting trace in the user. The present work takes exactly this step: it treats the affect evoked during agent interaction not merely as a UX variable but as an instance of neuroplastic conditioning.

2.2. Neuroplasticity and Experience-Dependent Change

The mechanistic basis of the framework is the well-established neuroscience of activity-dependent synaptic plasticity. The principle that co-active neurons strengthen their connections [9], realised in the mechanisms of long-term potentiation and long-term depression [10, 11, 12, 21], provides the account of how repeated experience physically reshapes neural pathways. That such reshaping follows sustained behavioural practice is documented by studies of experience-dependent structural change, from the hippocampal changes accompanying spatial expertise [22] to grey-matter changes following motor skill acquisition [23]. This literature establishes the substrate on which the present framework rests; our contribution is not to the neuroscience itself but to identifying a novel, ubiquitous, and previously unexamined setting

— everyday agent interaction — in which these mechanisms are routinely engaged.

2.3. Observation-Based Emotion Regulation

The practice at the centre of the framework belongs to a family of observation-based approaches to emotion regulation with substantial clinical support. Mindfulness-based cognitive therapy cultivates a decentred relationship to thoughts and feelings, treating them as passing mental events rather than facts, and is associated with reduced relapse in recurrent depression [24, 25, 26]. Metacognitive therapy directs attention to the process of thinking rather than its content [27]. The broader emotion-regulation literature characterises the top-down, prefrontally mediated modulation of affective responses and its role in psychological health [28, 29, 30, 31]. That sustained contemplative practice of this kind produces measurable structural and functional change in the brain is itself documented [32, 33].

These practices are, however, almost always cultivated in dedicated settings — formal meditation, therapy sessions, structured daily exercises — and their benefit in everyday life depends on transferring a capacity built in those settings to the moments where reactive patterns actually fire. The present framework differs not in the nature of the observation it proposes but in its setting: it locates the practice within an activity people already perform many times a day, at the exact moments reactive patterns are triggered, so that the reactive moment and the opportunity to observe it coincide.

2.4. Positioning

The contribution of this paper lies in the connection of these three literatures, which have developed independently. The HCI literature establishes that agent interaction reliably evokes affect but treats that affect as transient UX friction; the neuroscience literature establishes that repeated affective responses physically reshape neural pathways but has not examined agent interaction as a setting in which this occurs; and the emotion-regulation literature establishes that observation can intercept reactive responses but situates the practice outside the moments of ordinary technology use. Placing these together yields the paper's central claim: that the human–AI interaction loop is a high-frequency, naturally occurring site of neuroplastic conditioning, at which the same observation-based mechanism studied in clinical settings can be practised — for good or ill depending on how the loop is engaged. To our knowledge, no prior work has characterised everyday agent interaction

in these terms, nor proposed the interaction loop itself as a site for upstream neuroplastic intervention.

Table 1 summarises this positioning across the dimensions that define the present framework. Each strand of prior work supports some of these dimensions but none combines them; in particular, no prior work treats the interaction loop itself as the intervention site or locates the practice within ordinary, tool-native use.

Table 1: Comparison with representative related work across the dimensions of the proposed framework. ✓ = supported; ✗ = not supported; ∼ = partially supported.

Work	Human–AI interaction focus	Interaction loop as intervention site	Neuroplastic framing (LTP/LTD)	Feeling-tone gap (upstream)	Behind-the-scenes observation / three layers	Everyday / no-app setting
Reeves & Nass [17]	✓	✗	✗	✗	✗	✓
Picard [18]	✓	✗	✗	✗	✗	∼
Klein et al. [19]	✓	✗	✗	✗	✗	∼
Amershi et al. [20]	✓	✗	✗	✗	✗	✓
Draganski et al. [23]	✗	✗	✓	✗	✗	✗
Teasdale et al. [24]	✗	✗	∼	∼	∼	✗
Segal et al. [25]	✗	✗	∼	∼	∼	✗
Wells [27]	✗	✗	✗	∼	∼	✗
Gross [30]	✗	✗	∼	✓	✗	✗
Hölzel et al. [32]	✗	✗	✓	∼	∼	✗
This work (proposed)	✓	✓	✓	✓	✓	✓

3. Theoretical Foundations

This section develops the end-to-end account of conditioned reactive patterns that the remainder of the paper builds on. It is deliberately presented as background, using everyday real-world examples rather than the human–AI interaction setting that is the subject of this work; that setting is taken up in Section 4, once the mechanism is in place. We proceed from the smallest unit — the single neurone and its synapse — to the neurone path, to the deeply encoded paths that constitute implicit beliefs, to the millisecond-scale affective routing that activates them below awareness, and finally to the opposing synaptic mechanisms of long-term potentiation and long-term depression that respectively deepen and dissolve them. The argument is cumulative: each subsection supplies a component that the later sections require.

3.1. The Neurone Path: The Physical Substrate of Psychological Experience

The foundational claim is that psychological experience is physical. A thought, a habit, a fear, a belief, or a skill is not an abstract entity residing apart from the brain; it is the activity of a specific configuration of connected neurons [9]. Two neurons that fire in close temporal succession strengthen the synaptic connection between them, so that subsequent activation of the first more readily activates the second [10, 11]. Repeated across chains of neurons, this process produces a *neurone path*: a physical sequence of connected neurons that, once activated at its origin, propagates activation along its length to a characteristic outcome (Figure 2).

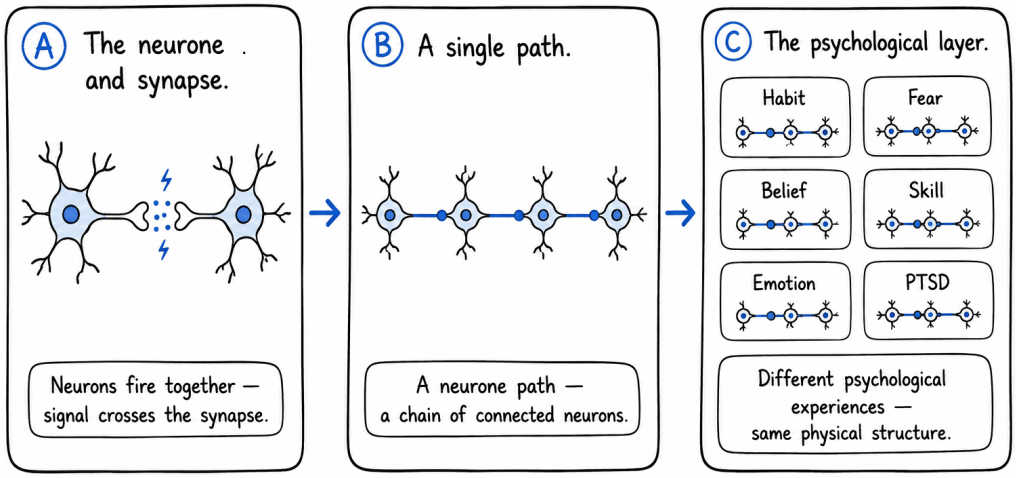

Figure 2: What neurone paths are. A single neurone communicates with the next by releasing neurotransmitters across the synapse (left). A chain of such connected neurons forms a neurone path (centre). Psychological phenomena as diverse as habits, fears, beliefs, skills, emotions, and post-traumatic stress share the same underlying physical structure — they differ in content and encoding depth, not in kind (right).

The psychological significance of a neurone path lies in what it does when it fires. A path activated by a particular situation routes toward a characteristic experience and behaviour. Critically, this routing occurs whether or not the person chooses it: once the path is activated at its origin, propagation is automatic. The subjective experience — the thought, the feeling, the urge — is the activation of the path, not a separate event that follows it (Figure 3).

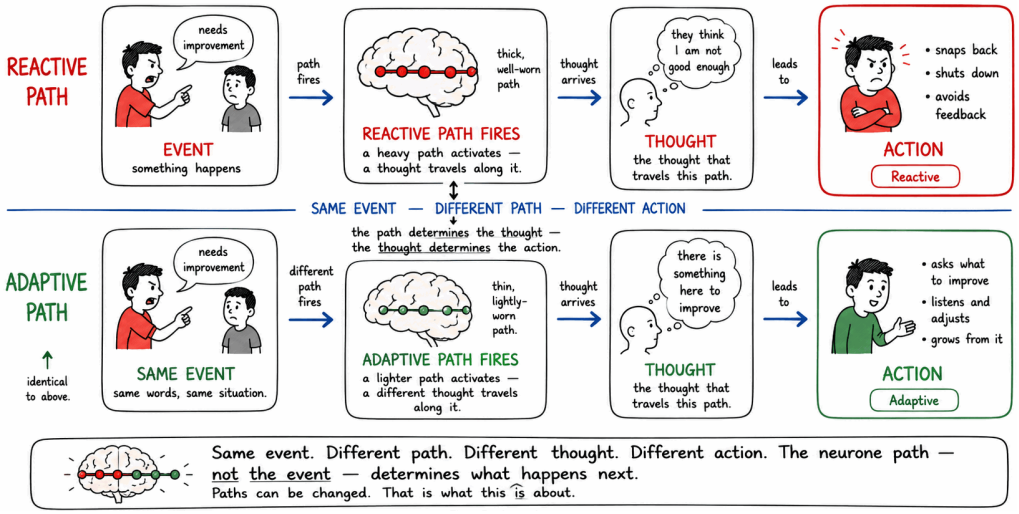

Figure 3: How a neurone path works, from event to action. The same event can activate different paths in different individuals, and the activated path — not the event itself — determines the resulting thought and action. A heavily encoded path (top) routes rapidly to a reactive outcome; a lighter path (bottom) routes to a considered response. The path determines the thought; the thought determines the action.

This reframing has a direct practical consequence. If a reactive pattern is a physical path, then changing the pattern is a matter of changing the path — not of willpower, insight alone, or the suppression of an unwanted output. The remainder of this section establishes how paths form, how they deepen, and how they can be dissolved.

3.2. How Neurone Paths Form

Neurone paths are encoded through three mechanisms. The first is *repetition*: a situation and a response that co-occur repeatedly strengthen their connection incrementally, each repetition adding a small increment of synaptic efficacy [9, 10]. A child told repeatedly that their efforts are not good enough encodes an inadequacy path gradually, over years, until it activates automatically in adulthood whenever their work is judged — no single episode is decisive, but each instance adds a small increment until the path fires on its own. The second is *single-event intensity*: under conditions of extreme arousal, stress hormones accelerate synaptic consolidation such that a single event can produce the encoding that would otherwise require thousands of repetitions [34, 35]. The nurse who experiences one overwhelming mass-casualty night encodes a danger path in a single exposure. The third

is *genetic and epigenetic inheritance*: predispositions and the consequences of ancestral experience can be transmitted, biasing the ease with which particular paths form [36] (Figure 4).

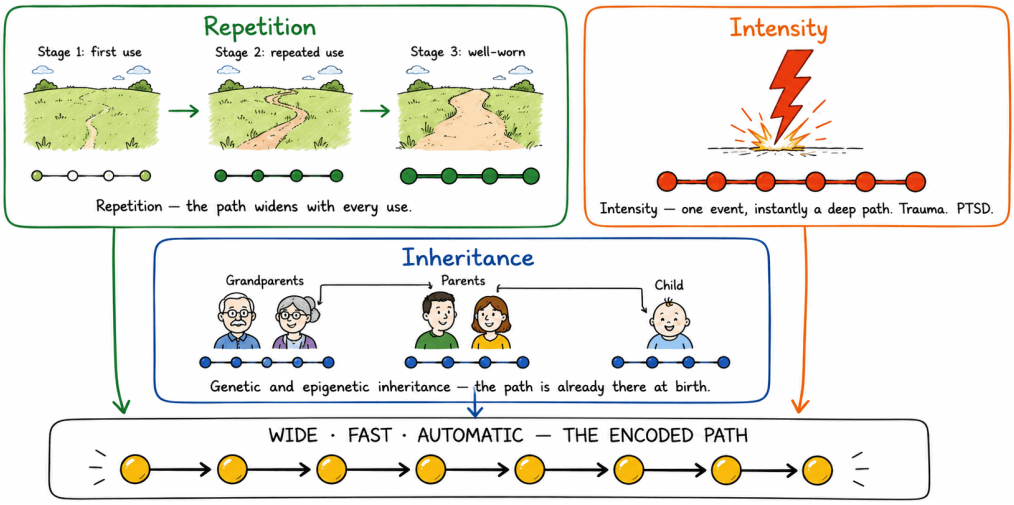

Figure 4: How neurone paths form. Three mechanisms encode paths into the brain's physical architecture: repetition (many small increments of synaptic strengthening over time), single-event intensity (one extreme event producing maximal encoding through stress-hormone-accelerated consolidation), and genetic or epigenetic inheritance (transmitted predispositions that bias path formation).

These mechanisms are not mutually exclusive; a given path may be inherited as a predisposition, established through repetition, and then intensified by a single event. What matters is that all three produce the same kind of physical structure, and all three are subject to the same mechanisms of deepening and dissolution described below.

3.3. Not All Paths Are Problems

Neurone paths are not pathological in general. The overwhelming majority of encoded paths are adaptive or neutral: the skill of reading, the ability to ride a bicycle, the recognition of a loved one's face, and the capacity for language are all deeply encoded paths, and their automaticity is precisely what makes them useful. Deep encoding is not the problem; it is the basis of all learning and competence (Figure 5).

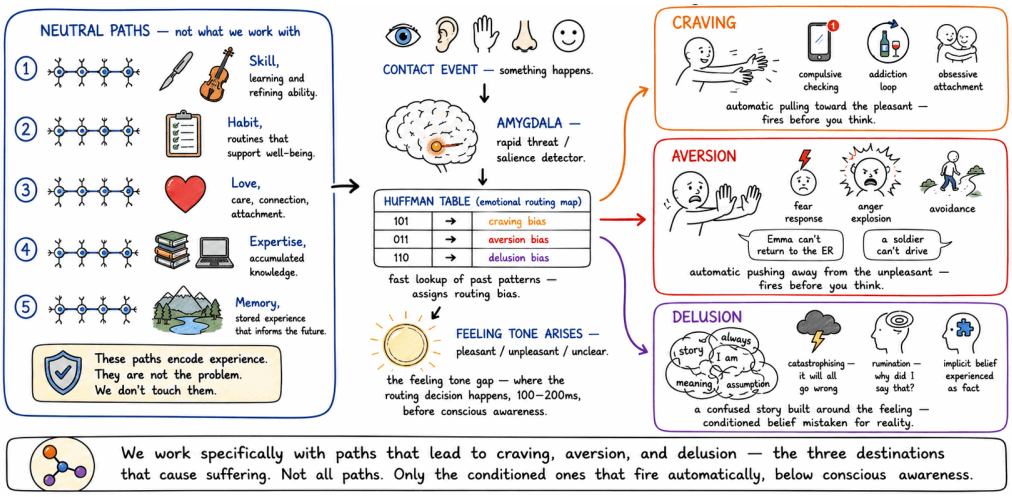

Figure 5: Not all paths are problems. The same physical mechanism of encoding underlies adaptive paths (skills, knowledge, recognition) and maladaptive reactive paths. Only paths that route automatically to reactive suffering — disproportionate, involuntary, and unwanted responses — are targets for intervention. Adaptive encoding is preserved; only the specific reactive paths that generate suffering are targeted.

The target of intervention is a specific subset: paths that route automatically to reactive suffering — responses that are disproportionate to the situation, involuntary, and unwanted by the person who experiences them. This scoping matters both practically and ethically: the aim is not to flatten affect or erase memory, but to dissolve the specific paths that produce reactions the person themselves wishes to be free of.

3.4. What Is a Situation? Situations Reach Us Through the Senses

Before describing how situations are encoded and matched, we must be precise about what a situation is. We propose that a situation is not an abstract whole but the specific set of sensory inputs through which an event reaches the brain. These inputs arrive through a closed set of channels: five *exteroceptive* channels that register the external world — sight, hearing, smell, taste, and touch — and at least one *internal* channel, memory, which registers self-generated content such as a surfacing recollection or a recalled voice. A felt inner state, such as the sense of being alone or unsupported, functions as a further internal channel. This closed-set formulation is deliberate: it converts an otherwise arbitrary list of situational features into a principled enumeration of the channels through which any situation can be

registered at all (Figure 6).

A clarification on terminology is warranted here. The brain does not index or retrieve whole "situations" as discrete objects; mechanistically, what propagates and is matched is a *pattern of sensory signals*, appraised rapidly and approximately by the amygdala and associated circuits against encoded patterns [7, 8]. We use "situation" throughout as a deliberate simplifying abstraction — a convenient name for that pattern of co-occurring sensory signals — because it makes the account legible without altering its mechanistic content. When we speak of a situation being matched to a path, we mean that the pattern of sensory signals characterising that situation is what the matching operates on.

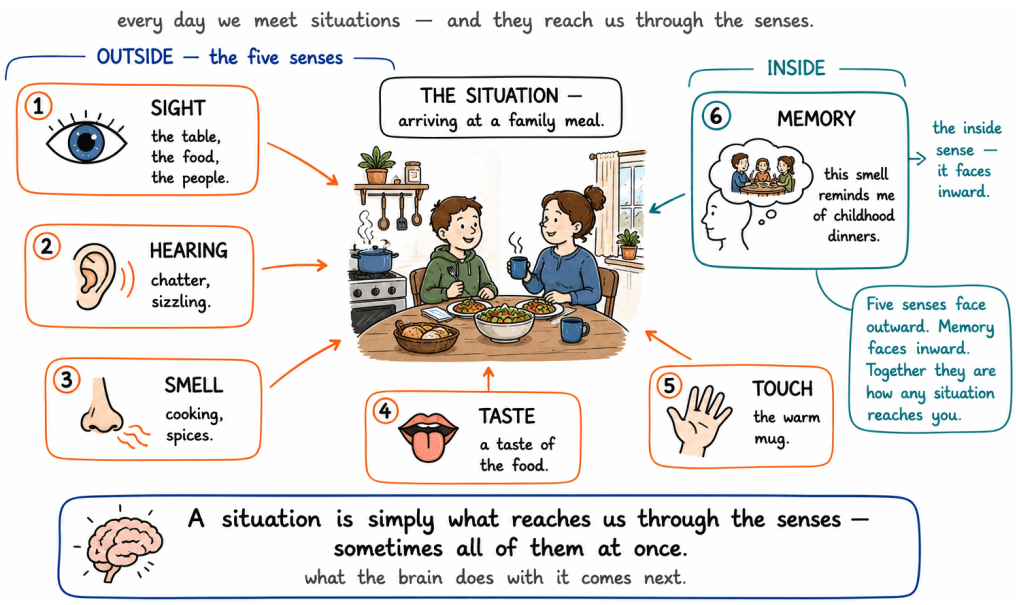

Figure 6: What a situation is. An everyday situation — here, arriving at a family meal — reaches the brain through a closed set of sensory channels: the five exteroceptive senses (sight, hearing, smell, taste, touch) and the internal channel of memory, which faces inward and can supply content with no external input at all. A situation is constituted by whichever channels are active, integrated together into a single recognised event.

Two consequences follow. First, a situation is composed only of the channels that are actually active; different situations recruit different subsets, and no situation requires all channels at once. Second, because memory is a genuine internal channel, a situation can be assembled partly or wholly from internally generated content — which is why a recollection alone, with the

external world entirely quiet, can constitute a situation capable of activating a path. This internal channel is central to the account of generalisation developed below. Our formulation here is a sensory-channel model grounded in multimodal integration; it is worth noting, as a historical antecedent rather than a foundation, that classical Buddhist analysis likewise treated the mind as a sixth sense base — alongside the five physical senses — whose objects are mental phenomena including memories [37, 38]. We do not develop this convergence here.

3.5. A Situation Links to a Neurone Path

A situation, once registered through its active channels, is matched against the brain's store of encoded paths. When the pattern of sensory inputs — appraised rapidly by the amygdala — matches the signature of an encoded path, that path is activated. The situation is thus the *key*, and the neurone path is what the key unlocks. Crucially, it is the pattern of sensory channels — not the event in the abstract — that performs the matching, and it is the activated path, not the event, that determines the resulting experience and behaviour (Figure 7).

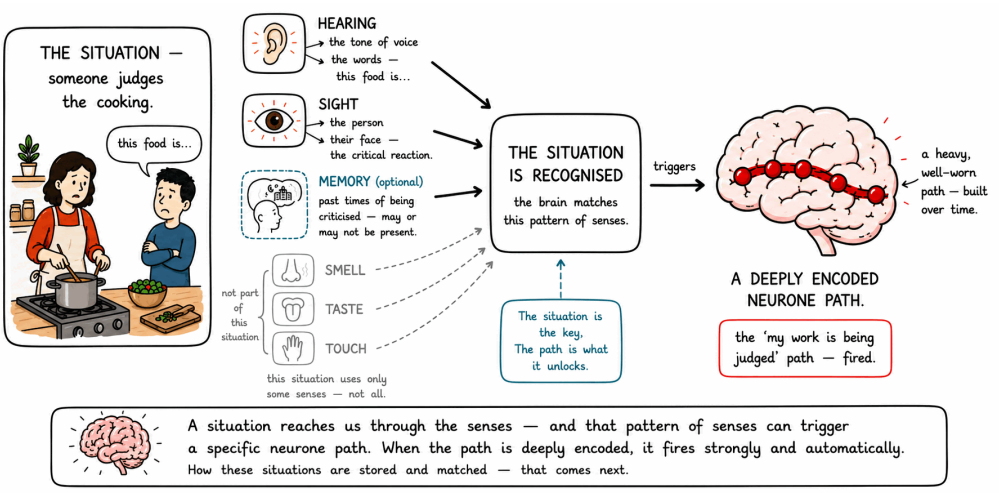

Figure 7: A situation links to a neurone path. A specific situation — someone commenting on a cook's food — reaches the brain through only the channels that carry it: hearing (the tone and the words) and sight (the person and their facial reaction), with memory as an optional internal contributor. The other senses are not part of this situation. This pattern of active channels matches, and thereby fires, a deeply encoded neurone path — here, the implicit-belief path "my work is being judged."

This figure also illustrates the point made above: only some channels carry a given situation. The cook's situation is composed of hearing and sight; smell, taste, and touch play no part in it, even though food is present. The situational key is therefore composed of only the channels that carried the path's formative attributes — a subset of the closed channel set, not the whole set. This subset structure is what makes the encoding model that follows both precise and tractable.

3.6. Deeply Encoded Paths as Implicit Beliefs: A Situation-Keyed Encoding Model

The most deeply encoded reactive paths function as *implicit beliefs*: automatic routing assumptions that determine how a situation is interpreted before any deliberate reasoning occurs. A person does not consciously decide that criticism means they are inadequate; the felt sense of inadequacy arrives first, in milliseconds, and conscious reasoning proceeds from it [39, 40]. The implicit belief is not a proposition the person holds; it is the path that fires.

We propose that this routing operates as a situation-keyed compressed encoding system, analogous to Huffman encoding in information theory [41]. In Huffman encoding, the most frequently occurring symbols are assigned the shortest codes, minimising the average length of the encoded message. In the brain's affective architecture, the most frequently or intensely encoded paths become the shortest, fastest, most automatic routes: they fire first, at the lowest stimulus intensity, and below conscious awareness. The *key* to this lookup is the situation, decomposed — as established in Section 3.4 — into sense-tagged attributes across the closed channel set: what is heard (tone, words), seen (face, person, place), smelled, and internally registered (memory, inner state). The amygdala performs approximate pattern matching of these incoming attributes against encoded path signatures, and returns a pre-cognitive affective routing decision — a feeling tone — in under 200 milliseconds [7, 8] (Figure 8).

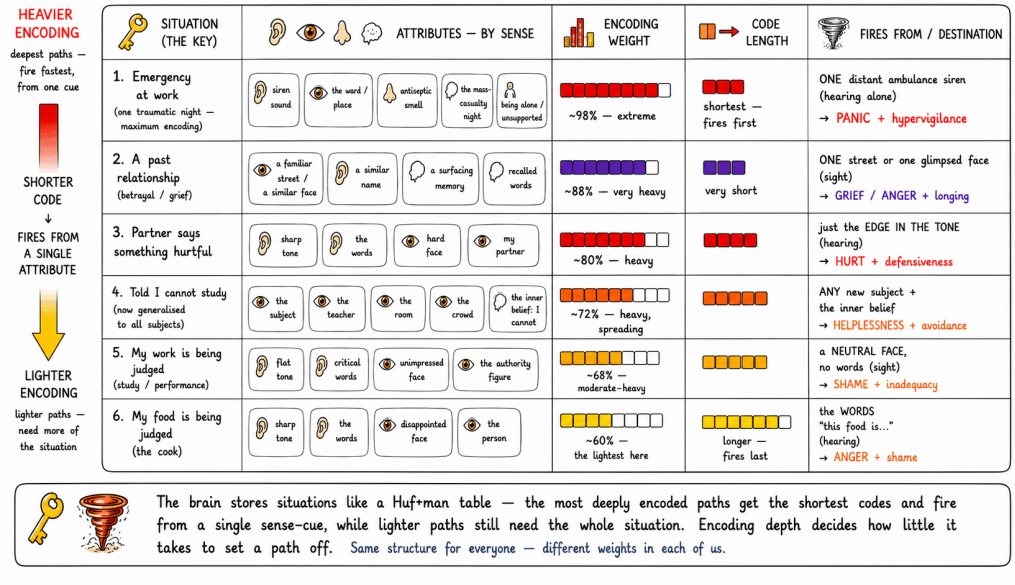

Figure 8: The situation-keyed encoding model. Each row represents a conditioned reactive path indexed by a situational key comprising attributes such as tone, facial expression, words, location, sound, and memory. Deeply encoded paths (top rows) carry the greatest weight and fire fastest and most automatically, analogous to the shortest codes in Huffman encoding. Incoming situations are matched approximately against these keys to generate a pre-cognitive affective routing decision in under 200 milliseconds.

This model extends existing accounts of rapid affective appraisal [7, 8, 40] by specifying the structure of the situational key and the compression principle that determines which paths dominate. It also sets up the central phenomenon addressed next: the behaviour of these keys as encoding depth increases.

3.7. Encoding-Depth-Dependent Stimulus Generalisation: The Tornado Effect

We propose that the situational specificity required to activate a conditioned path is inversely proportional to its encoding depth. A thinly encoded path requires the full complement of original situational attributes to fire: the tone, the face, the words, and the location must all be present. As encoding depth increases, progressively fewer attributes are required, until a heavily encoded path fires from a single matching attribute. We term this *encoding-depth-dependent stimulus generalisation*, and offer the metaphor of a tornado to make it intuitive: a light breeze is felt only on direct contact, whereas a

tornado exerts its pull far beyond its visible boundary [42]. The deeper the encoding, the wider the path's pull radius (Figure 9).

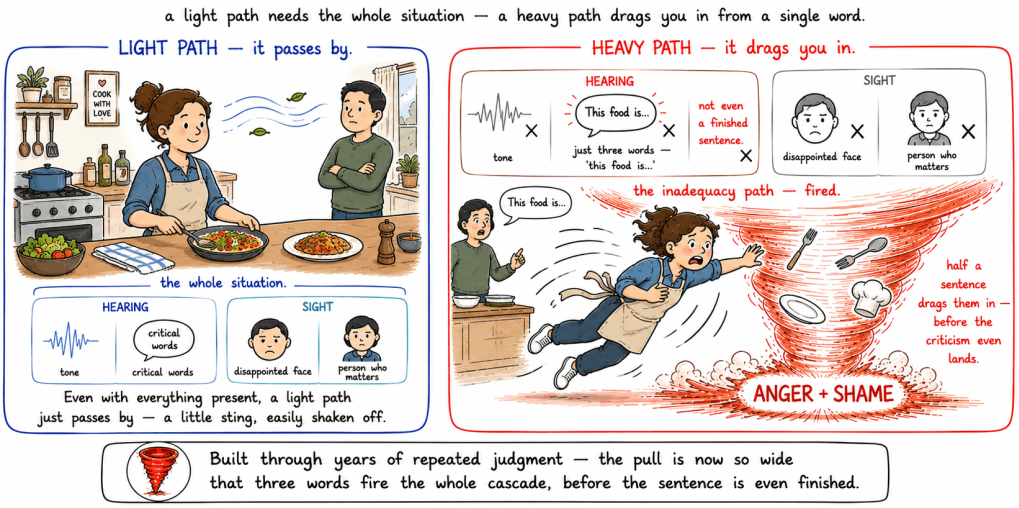

Figure 9: The tornado effect in a repetition-encoded path (cook scenario). A normally encoded path requires all situational attributes — tone, facial expression, the specific person, and critical words — to be present to fire (left). After deep encoding through years of repeated criticism, the path fires from a single attribute: the words "this food is..." alone activate the full reactive cascade, regardless of tone, speaker, or setting (right).

The same principle explains the hallmark feature of trauma: activation from stimuli that share almost nothing with the original event. When a path is encoded at maximal intensity in a single event, its pull radius becomes extremely wide, and a single peripheral attribute — a sound, a smell — suffices to fire the entire cascade [34, 35] (Figure 10).

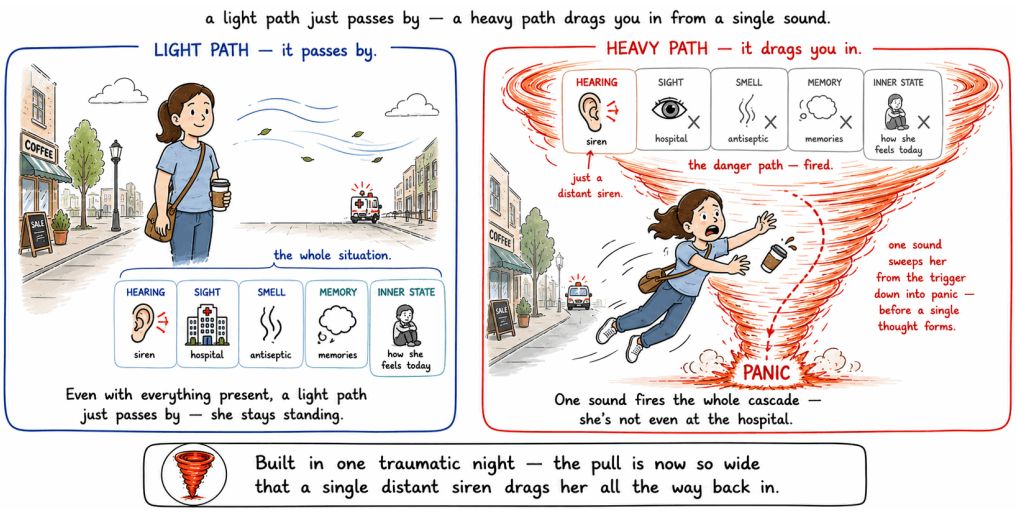

Figure 10: The tornado effect in an intensity-encoded path (hospital scenario). Following a single mass-casualty event encoded at maximal intensity, the danger path acquires an extremely wide pull radius. A distant ambulance siren — a single attribute, absent the original people, place, and circumstances — is sufficient to activate the full reactive cascade. This is the mechanistic basis of trauma responses triggering from apparently unrelated stimuli.

Encoding-depth-dependent stimulus generalisation reframes a puzzling observation as a direct consequence of encoding depth, and predicts that reducing encoding depth — narrowing the path — should correspondingly narrow the pull radius, restoring the requirement for more complete situational matching before activation.

3.8. Affective Routing and the Feeling Tone Gap

We now specify the temporal structure of activation, because it is here that intervention becomes possible. Following situational registration, the amygdala generates a pre-cognitive affective signal — a bare feeling tone of pleasantness, unpleasantness, or neutrality — within approximately 100–200ms, prior to full prefrontal engagement [7, 8, 39]. This feeling tone is not yet an emotion, a thought, or a reaction; it is the bare affective valence that precedes and shapes them. The constructed emotional experience — the narrative of what is happening and what it means — is assembled afterward [40] (Figure 11).

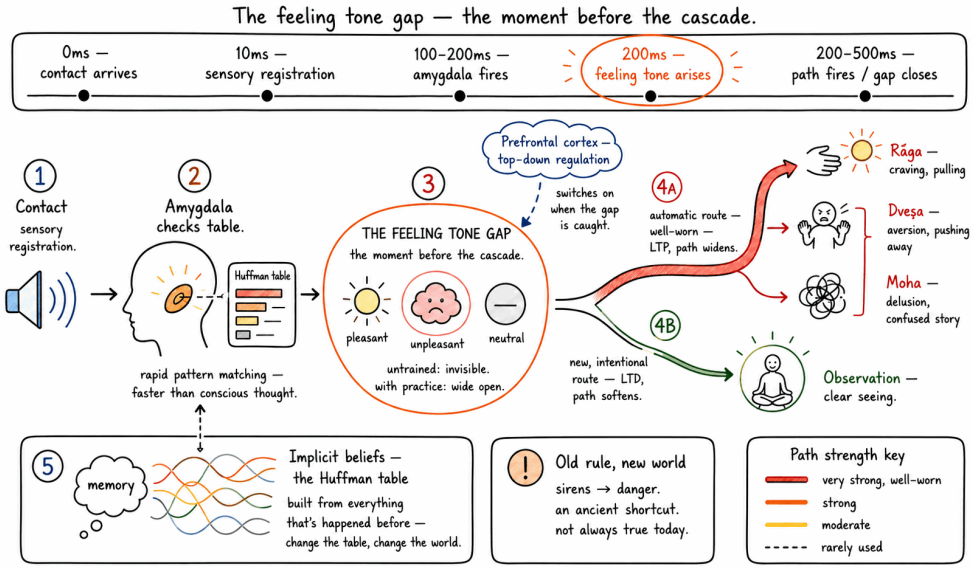

Figure 11: How the brain routes thoughts, and the feeling tone gap. A situation is registered and matched against encoded paths; the amygdala returns a pre-cognitive feeling tone within 100–200ms; and only afterward is the conscious thought and reactive behaviour constructed. The brief prefrontal-amygdala regulatory window between the feeling tone and the completion of the reactive cascade — the feeling tone gap — is the sole point at which the cascade can be intercepted upstream.

Between the arising of the feeling tone and the completion of the reactive cascade lies a brief prefrontal-amygdala regulatory window during which top-down modulation can intercept the cascade before it completes [29, 28, 30, 31]. We call this window the *feeling tone gap*. In untrained awareness it is effectively invisible: the feeling tone and the reaction appear as a single seamless event. The central proposal is that this gap is the only point at which upstream intervention — and therefore structural dissolution — is possible.

3.9. How Paths Deepen: Long-Term Potentiation

Each time a reactive path fires to completion, long-term potentiation operates on the synaptic connections along its length: AMPA receptors are inserted into the post-synaptic membrane, and the path is strengthened [10, 11, 12]. The path becomes wider, its activation threshold lower, its firing faster and more automatic, and its pull radius — following the tornado principle — wider. This is why reactive patterns worsen with repetition rather

than exhausting themselves, and why avoidance, far from relieving a fear, reliably intensifies it: each avoidance episode fires and therefore strengthens the very path it seeks to escape [43, 44] (Figure 12).

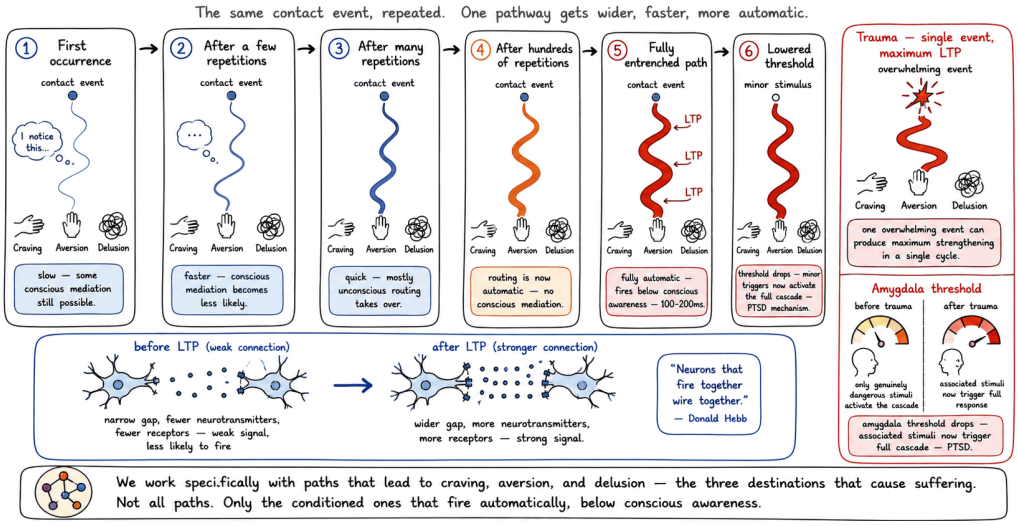

Figure 12: How paths deepen through long-term potentiation. Each completed firing of a reactive path strengthens its synaptic connections, widening the path, lowering its activation threshold, and increasing its firing speed and automaticity. Untreated reactive patterns deepen over time because every triggered episode functions as a strengthening event.

We illustrate path deepening in two contrasting cases, presented separately because they represent the two principal formation mechanisms. The first is the cook, whose inadequacy path was built by *repetition*. No single episode of criticism was decisive; instead, each instance added a small increment of long-term potentiation, and over years these increments accumulated into a deeply encoded path. Every subsequent completed firing — each adult episode in which criticism of the cook's food fires the path to a full reaction — continues this deepening, widening the path and lowering its threshold further (Figure 13).

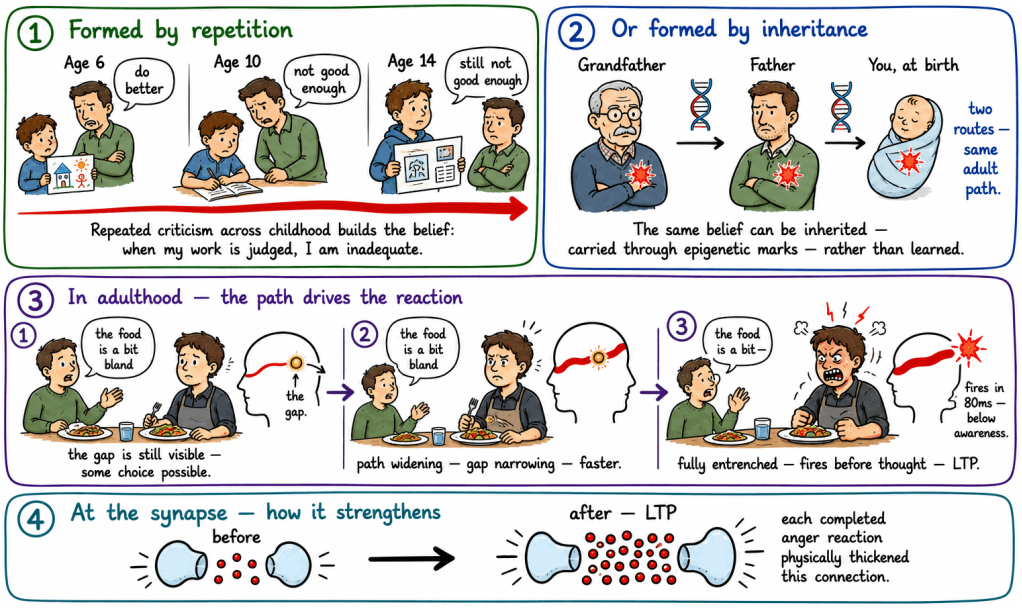

Figure 13: Path deepening through repetition (the cook). The inadequacy path is built incrementally: each episode of criticism adds a small increment of long-term potentiation, and over years these accumulate into a deeply encoded path. Every completed firing thereafter deepens it further, widening the path and lowering its activation threshold.

The second is the hospital case (Emma), whose danger path was built by *single-event intensity*. Here one overwhelming night produced, in a single exposure, the depth of encoding that repetition would take years to achieve — because extreme physiological arousal accelerates synaptic consolidation [34, 35]. Thereafter every triggered episode of the panic cascade operates as a further potentiation event, so that untreated the path deepens over time rather than fading (Figure 14).

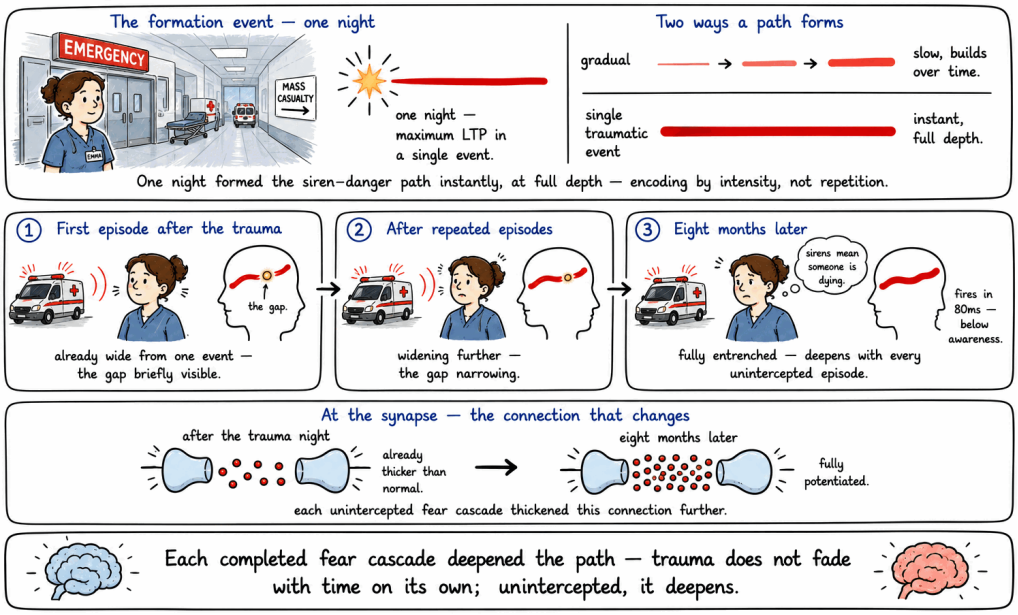

Figure 14: Path deepening through single-event intensity (the hospital case, Emma). A single night of extreme arousal produces, in one exposure, the encoding depth that repetition would require years to build. Each subsequent triggered episode of the panic cascade acts as a further potentiation event, deepening the path rather than exhausting it.

The two cases arrive at the same endpoint — a deeply encoded reactive path — by different temporal routes: many small increments in the cook, one large increment in Emma. The implication is identical for both: any intervention which allows the path to continue firing to completion — however it manages the aftermath — leaves the deepening mechanism intact. Structural change requires interrupting the firing itself.

3.10. Which Paths Require Rewiring

Because most encoded paths are adaptive, a principled criterion is needed for identifying intervention targets. A path is a candidate for rewiring when its activation is disproportionate to the situation, involuntary, and productive of suffering the person wishes to be free of. Adaptive paths — skills, knowledge, meaningful emotional responses — are explicitly preserved (Figure 15).

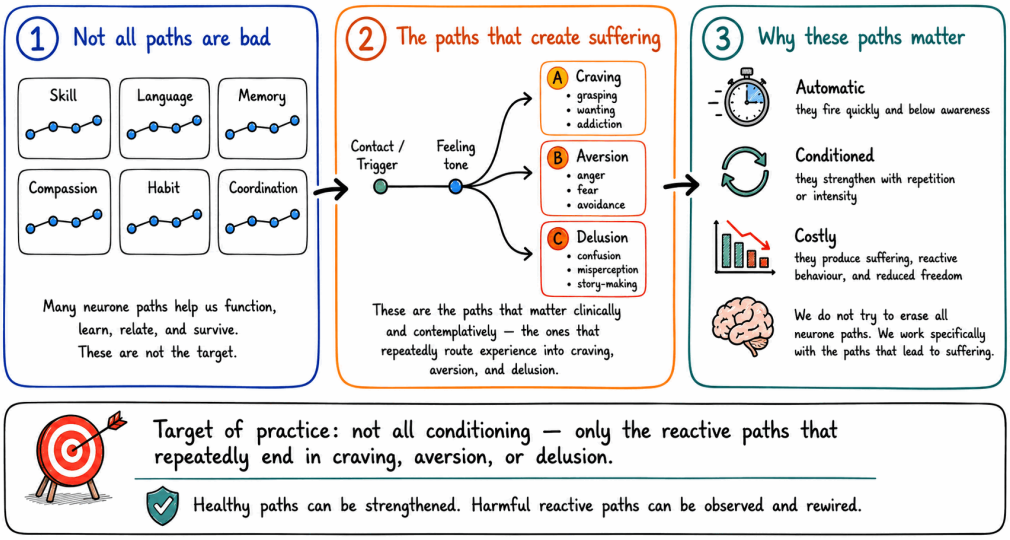

Figure 15: Which neurone paths require rewiring. Intervention targets are restricted to paths that route automatically to reactive suffering: disproportionate, involuntary, and unwanted responses. Adaptive and neutral paths are preserved. Identifying the specific target path — its situational key and reactive destination — is the first step of any intervention built on this model.

Identifying the specific target path, including its situational key and its characteristic reactive destination, is the first practical task for any intervention grounded in this model.

3.11. How Paths Dissolve: Long-Term Depression Through Behind-the-Scenes Observation

Long-term depression is the synaptic mechanism opposite to potentiation: under conditions of activation without completion, AMPA receptors are internalised from the post-synaptic membrane, phosphatases activate, and the synaptic connection physically weakens [11, 12, 21, 45]. The central mechanistic claim is that engaging the feeling tone gap — observing the path as it begins to fire, rather than allowing it to complete — creates precisely these conditions. Each interception produces a small increment of LTD. Individually these increments are negligible; accumulated across hundreds of sessions, they produce measurable structural reorganisation, narrowing the path, raising its threshold, and reducing its pull radius.

We term the practice that achieves this interception *behind-the-scenes observation*, and we are precise about what it is and is not. It is not calming the

feeling, reframing it, replacing it with a more positive thought, or accepting it. It is, specifically, the act of consciously watching the neural machinery operate in real time as the reaction begins: registering that the amygdala has matched the situation and fired, that a conditioned path is beginning to route the signal toward its characteristic reactive destination, and that if the cascade completes the path will deepen through potentiation [46]. The content of the observation is mechanistic rather than emotional — in the first person, "a path is firing here; it was built through past encoding; if it completes, the encoding deepens; I am watching this happen; I am not the path but its observer." We propose that it is this mechanistic quality that distinguishes the practice from general awareness and explains its synaptic effect: watching one's own reactive architecture operate constitutes an unusually deep form of metacognitive engagement, which recruits the strongest top-down prefrontal–amygdala regulatory signal and thereby the most complete interruption of the cascade before completion [29, 28, 30]. In this precise sense, awareness of the mechanism is itself the mechanism: it is the seeing of the path, rather than any effort applied after the feeling, that creates the activation-without-completion under which LTD operates.

We decompose the practice into three layers of observation that trace the mechanism's own stages in the order they occur — signal, feeling tone, and neural routing. These layers serve both as a description of what is observed and, for people in whom the reactive cascade fires too rapidly to be observed directly, as a graded learning scaffold approached one stage at a time. In the first layer the person notices the signal arriving — registering that something has been received through one of the six sense bases (sight, hearing, smell, taste, touch, or the internal sense of memory), including the case in which the signal is an internally arising recollection rather than an external event. In the second layer they notice the bare feeling tone the signal evokes — the pre-cognitive valence of pleasantness, unpleasantness, or neutrality that arises before any emotion or narrative ("there is unpleasantness here") [15]. In the third layer they see the behind-the-scenes neural process itself in motion: that the amygdala has matched the signal and found a path, that a signal is now routing along that conditioned path toward its reactive destination — craving, aversion, or delusion — and that if the route completes, the path widens and the encoding deepens ("the amygdala has found a path; it is routing this toward aversion; if it completes, the encoding deepens; I am watching this happen; I am not the path but its observer"). The third layer is behind-the-scenes observation proper, and it is here that the interception

occurs; the first two layers bring the observer to it. With practice the three collapse into a single compressed act — registering the signal and its feeling tone and, in the same moment, seeing the neural process routing beneath them.

Whereas the three layers describe how deeply the process is seen, a single completed interception also has a temporal anatomy: the bare feeling tone is noticed as it arises, it is recognised as the activation of a path rather than a fact about the world, observation is sustained without allowing the cascade to complete, and the path is witnessed subsiding un-acted-upon [24, 26, 27, 25]. This gives the observation, so often described in general terms, a specific synaptic interpretation (Figure 16).

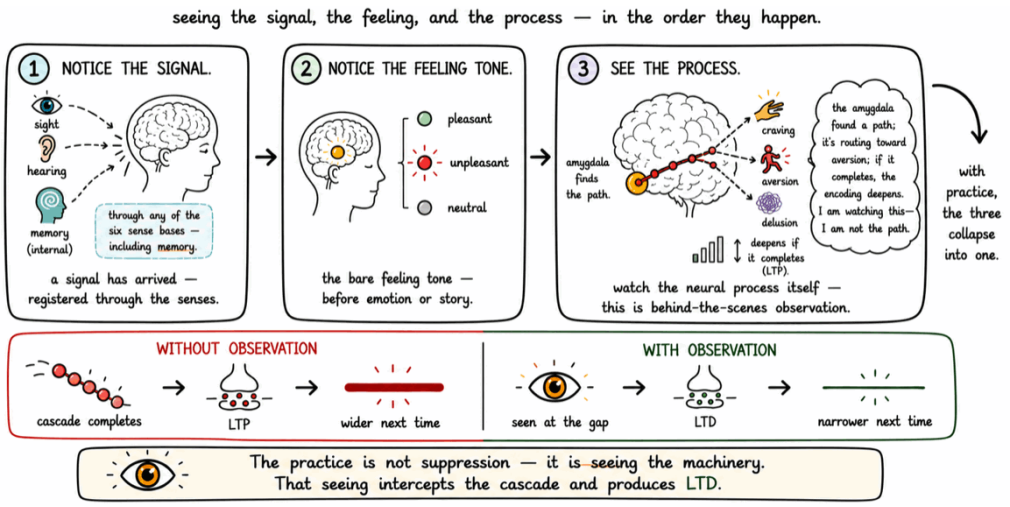

Figure 16: How paths dissolve through long-term depression, via the three layers of behind-the-scenes observation. The layers trace the mechanism's stages in the order they occur: noticing the signal arriving through the senses (Layer 1), noticing the bare feeling tone it evokes (Layer 2), and seeing the neural process itself in motion — the amygdala matching the signal and a path routing toward craving, aversion, or delusion, which deepens if the route completes (Layer 3). Seeing the process at Layer 3 intercepts the cascade before completion; each interception produces a small increment of long-term depression, weakening the path.

This mechanism defines the distinction between suppression and dissolution. Suppression manages the behavioural output of a cascade that has already fired to completion, leaving the path — and its potentiation — intact. Dissolution intercepts the cascade upstream, at the feeling tone gap, so

that the path does not fire to completion and long-term depression operates instead. Only the latter produces structural change in the path itself.

A single interception changes little; the significance of the mechanism is cumulative. Across repeated sessions, two processes run in parallel. The reactive path, no longer permitted to fire to completion, undergoes incremental long-term depression and progressively weakens. At the same time, the act of observation is itself a neural activity, and each time it is performed it is strengthened through the same Hebbian potentiation that builds any path — so the capacity to observe grows stronger with practice. Over many sessions these opposing trajectories cross: the observing capacity, initially weak, becomes stronger than the reactive path, which has narrowed. That this kind of repeated mental training produces measurable structural change in the brain is well established — experience-dependent grey-matter change follows sustained practice in domains from spatial navigation to juggling [22, 23], and contemplative training in particular is associated with structural and functional reorganisation in regions supporting attention and emotion regulation [32, 33]. Guided observation at the feeling tone gap engages precisely this capacity for structural reorganisation, directed specifically at the reactive path (Figure 17).

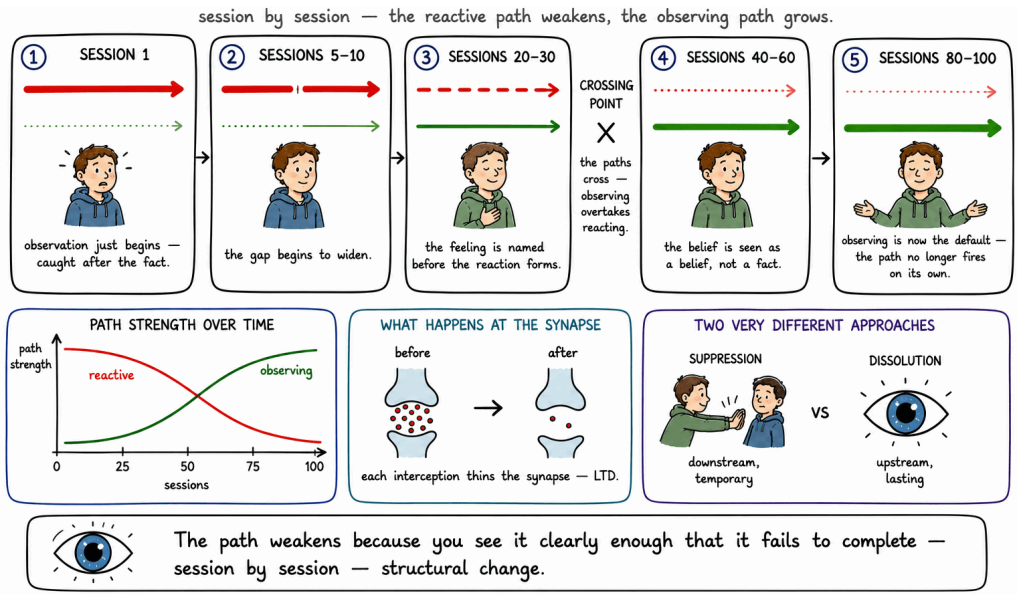

Figure 17: How observation rewires the brain across sessions. Repeated interception at the feeling tone gap drives two parallel trajectories: the reactive path weakens through accumulated long-term depression (its width and pull radius contracting session by session), while the observing capacity strengthens through potentiation each time it is exercised. Over many sessions the trajectories cross — the observing capacity overtakes the reactive path — after which the situation that once triggered the automatic cascade no longer does so. The structural change is cumulative and gradual, not a single event.

We illustrate this dissolution process in the same two cases used above. For the cook, repeated observation of the "my work is being judged" path at the feeling tone gap — catching it as it begins to fire rather than completing the reaction — accumulates long-term depression across sessions, until food criticism no longer fires the full cascade automatically (Figure 18).

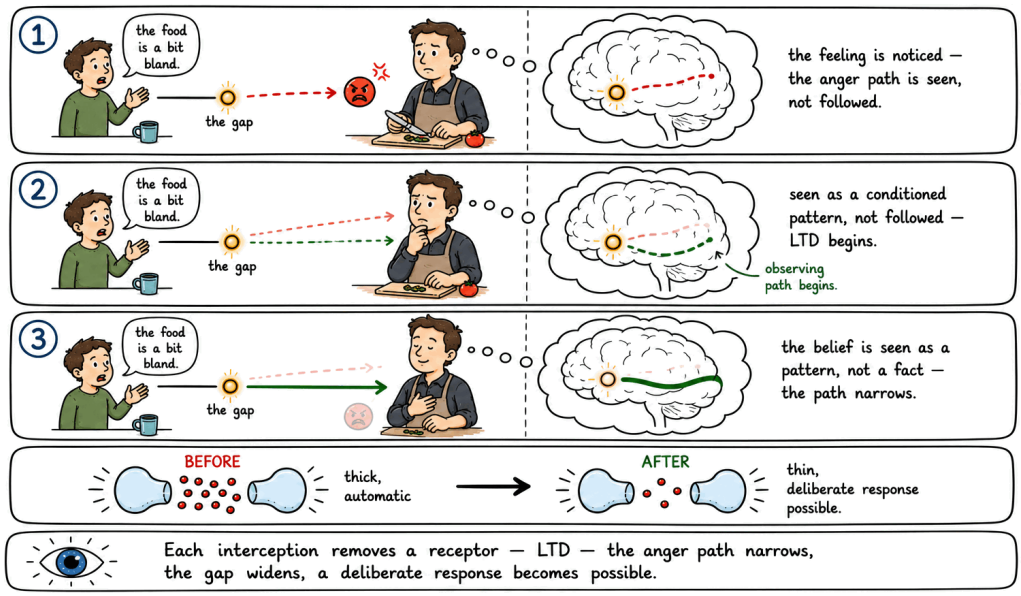

Figure 18: Path dissolution through observation (the cook). Repeated observation at the feeling tone gap — catching the "my work is being judged" path as it begins to fire, rather than completing the reaction — accumulates long-term depression across sessions. The path narrows, its threshold rises, and its pull radius contracts, until food criticism no longer fires the full reactive cascade automatically.

Because the mechanism is defined by interception at the gap rather than by how the path was originally encoded, it applies identically to the intensity-encoded case. For Emma, repeated observation of the danger path at the feeling tone gap — noticing the panic as it begins to arise, before the cascade completes — accumulates the same long-term depression, session by session, narrowing a path that was built in a single night (Figure 19).

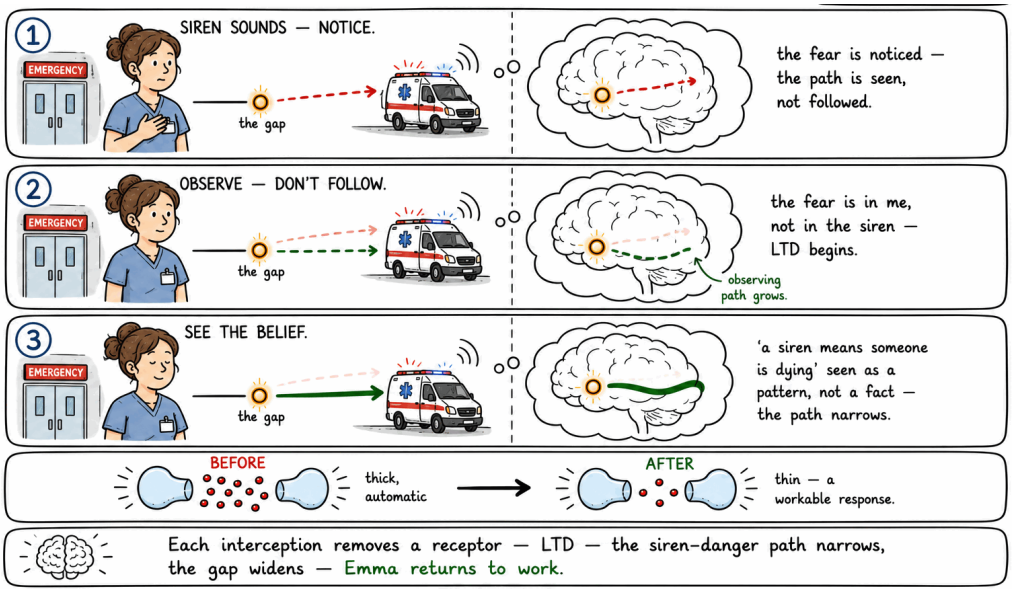

Figure 19: Path dissolution through observation (the hospital case, Emma). Though the danger path was encoded in a single intense event, it dissolves through the same mechanism as the cook's: repeated interception at the feeling tone gap accumulates long-term depression, narrowing the path and contracting its pull radius, until a siren no longer fires the full panic cascade. The encoding history differs; the dissolution mechanism is identical.

Taken together, the components of this section constitute an end-to-end model of conditioned reactive patterns: reactive patterns are physical paths; paths are encoded by repetition, intensity, and inheritance; a situation reaches the brain through a closed set of sensory channels and, as a situation-keyed implicit belief, activates a path whose pull radius widens with encoding depth; the path fires through a pre-cognitive feeling tone that opens a brief regulatory gap; potentiation deepens it when it fires to completion; and long-term depression dissolves it when it is intercepted at the gap through structured observation. Having established this mechanism with everyday examples, we now turn to the setting that is the subject of this paper: how these same paths are engaged, and how they may be rewired, during human–AI agent interaction (Section 4).

4. Reactive Paths in Human–AI Agent Environments

The mechanism developed in Section 3 was illustrated with everyday situations — a criticised cook, a nurse and a siren. This section argues that the same mechanism operates, largely unnoticed, in a setting that now occupies a growing share of daily life: interaction with AI agents [47]. We first characterise the structure of that interaction as an iterative loop, then show that each cycle of the loop is a contact event in the precise sense of Section 3.4, identify which conditioned paths it activates, and finally locate within the loop both the potentiation that deepens those paths and the feeling tone gap at which they can be intercepted.

4.1. The Structure of Agent Interaction: The Iterative Loop

Interaction with an AI agent — a conversational assistant, a coding copilot, or a generative tool for images, text, or design — proceeds through a characteristic iterative loop. The person issues a request; the agent returns a result; the person evaluates that result; and, on the basis of that evaluation, revises the request and issues it again. A prompt is written, an output appears, the output is judged, and a new prompt follows [6, 3]. This four-phase cycle — request, result, appraisal, revision — repeats many times within a single task and many times again across a working session (Figure 20).

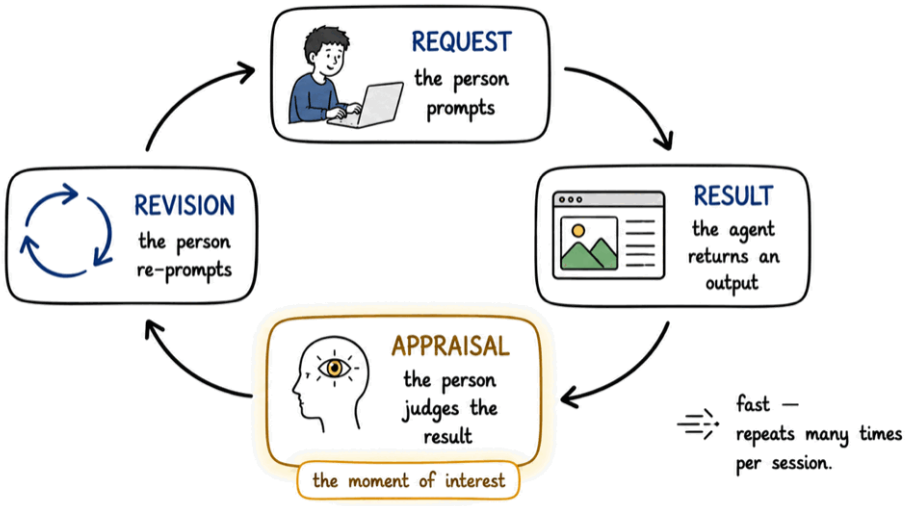

Figure 20: The iterative structure of human–AI agent interaction. The person issues a request; the agent returns a result; the person appraises the result; and the person revises and re-issues the request. This four-phase loop — request, result, appraisal, revision — repeats rapidly, many times within a single task. The appraisal phase, in which the returned result meets the person, is the point of interest for this paper.

Two properties of this loop distinguish it from most everyday situations in which reactive patterns are encountered, and both will matter for the argument. First, it is *high-frequency*: where the situations of Section 3 might recur a few times a day, the interaction loop can complete dozens of times within a single session, each completion delivering a fresh result to be appraised. Second, the interval between stimulus and response is *short and sharply bounded*: the result appears at a definite moment, and the reactive response — the revised prompt, often typed in evident frustration — follows within seconds. The loop thus compresses into minutes a volume of stimulus–response cycles that, in ordinary life, would be spread across days.

4.2. Each Cycle Is a Contact Event

The appraisal phase of the loop — the moment at which the returned result meets the person — is a contact event in the exact sense established in Section 3.4: a situation reaches the person through their sensory channels and is matched against encoded paths. When a generated image appears on screen, the situation arrives predominantly through sight; when a coding agent's output breaks a running program, it may arrive through sight together with an internal channel — the recalled history of similar failures.

The returned result is, in the terms of Section 3.5, a situational key: a pattern of sensory attributes that the amygdala matches, within approximately 100–200ms, against the person's store of encoded paths [7, 8] (Figure 21).

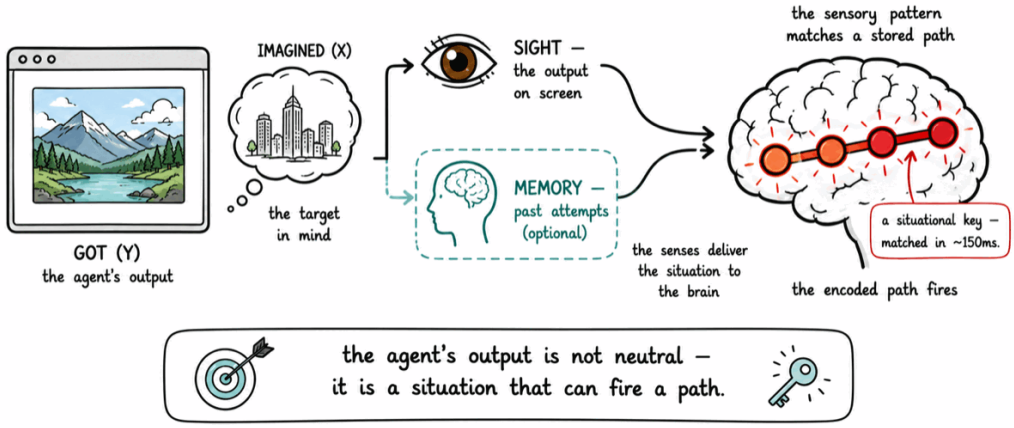

Figure 21: An agent's result as a contact event. When the agent returns a result — here, a generated image that does not match what was wanted — the result reaches the person through their senses (predominantly sight) and, optionally, through the internal channel of memory (the recalled history of previous attempts). This pattern of sensory attributes is a situational key, matched by the amygdala against encoded paths exactly as any other situation is. The agent's output is thus not a neutral stimulus but a situation capable of firing a conditioned reactive path.

This identification is the pivot of the paper. It means that the agent's output is not a neutral technical artefact but a situation in the full sense of Section 3 — and therefore that it can, and routinely does, activate conditioned reactive paths. Because the loop is high-frequency, it supplies these contact events in dense succession. Ordinary agent use is thus a continuous stream of situations, each capable of firing a path, arriving far faster than in most other domains of life.

4.3. Which Paths the Agent Fires

The paths that agent interaction activates are not new or agent-specific. They are the same conditioned paths a person already carries, now keyed by a novel class of situation. Two are especially common, and both were foreshadowed by the everyday examples of Section 3.

The first is the path that treats *one's own output as being judged*. When a person prompts an agent and receives a result that falls short, the situation

— one's effort, evaluated, found wanting — is structurally the same as the one that fires the cook's inadequacy path in Section 3.5. To a brain that carries a deeply encoded "my work is being judged" path, a disappointing generated image is not merely a technical outcome; it is an instance of the very situation the path was built to detect. The evaluation the person performs on the agent's output doubles, below awareness, as an evaluation felt to be performed on the person [48, 49].

The second is the path that concludes *one cannot do this*. When repeated prompts fail to produce the intended result, the situation — sustained effort at a task without success — can fire an encoded inadequacy or helplessness path of exactly the kind laid down, in Section 3.2, by a childhood of being told one's efforts are not good enough. The agent becomes the occasion for a much older reactive pattern.

Alongside these, agent interaction reliably evokes reactive paths of impatience (the demand that the result arrive now), of perfectionism (the refusal to accept any result short of an unspecified ideal), and of control (the felt need for the tool to comply exactly) [50]. In each case the reactive response completes in the familiar way: a feeling tone of unpleasantness, a constructed narrative ("it never understands," "I am wasting my time," "I can't even do this"), and a reactive behaviour — an abrupt, often harsher re-prompt (Figure 22).

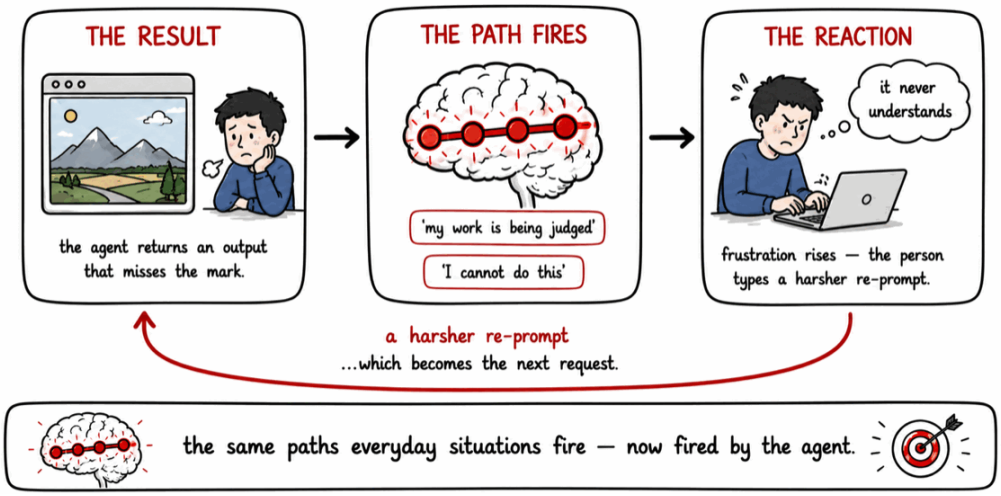

Figure 22: An agent's result firing a reactive path. A disappointing result (left) reaches the person as a situational key and matches a deeply encoded path — here, "my work is being judged" or "I cannot do this" (centre). The path fires to completion: a pre-cognitive feeling tone of unpleasantness, a constructed narrative, and a reactive behaviour — a harsher, more frustrated re-prompt (right). The same paths that everyday situations fire (Section 3) are fired here by the agent's output.

The reactive behaviour is worth dwelling on, because it closes the loop. The frustrated re-prompt is not outside the interaction; it is the next request, which produces the next result, which is appraised in turn. The completion of the reactive cascade is thus fed directly back into the loop as its next input.

4.4. The Loop as a Potentiation Engine

Section 3.9 established that each time a reactive path fires to completion, long-term potentiation strengthens it: the path widens, its threshold falls, and it fires faster and more automatically the next time [10, 11, 12]. The interaction loop, left unobserved, is therefore a potentiation engine. Each cycle in which a disappointing result fires the "my work is being judged" path to completion is one more strengthening event; and because the loop is high-frequency, these events accumulate far faster than in ordinary life. A single frustrating prompting session may complete the reactive cascade a dozen times or more, each completion depositing a small increment of potentiation on the same path (Figure 23).

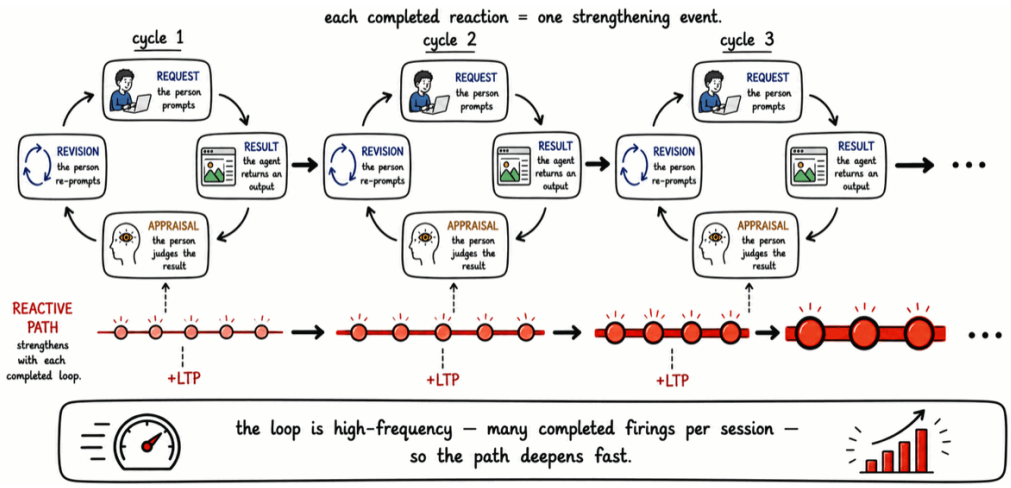

Figure 23: The interaction loop as a potentiation engine. Each cycle in which the agent's result fires a reactive path to completion deposits an increment of long-term potentiation, widening the path and lowering its threshold. Because the loop is high-frequency — completing many times within a single session — these increments accumulate rapidly. Repeated, unobserved agent interaction can therefore deepen a reactive path faster than most everyday situations do.

This yields an observation with some weight. Hundreds of millions of people now spend substantial time each day inside this loop [1, 2]. To the extent that the loop repeatedly fires reactive paths of impatience, perfectionism, and self-judgment to completion, everyday agent interaction may be quietly training precisely those patterns — not because the technology is designed to, but because the structure of the loop, combined with the ordinary human tendency to react to disappointing results, makes each session a dense sequence of potentiation events. The same property that makes the loop a training ground for reactive patterns, however, is what makes it available for the opposite purpose.

4.5. The Feeling Tone Gap Within the Loop

Section 3.8 located the sole point of upstream intervention at the feeling tone gap: the brief window between the arising of the pre-cognitive feeling tone and the completion of the reactive cascade. Within the interaction loop, this gap has an unusually definite location. It falls at the appraisal phase — in the moment between the result appearing and the reactive re-prompt being issued [51]. When a generated image resolves on screen and is not what was wanted, there is a brief interval, typically under a second,

in which a feeling tone of unpleasantness has arisen but the frustrated re-prompt has not yet been typed. That interval is the feeling tone gap, and in the interaction loop it is comparatively easy to locate, because the stimulus (the result appearing) and the response (the re-prompt) are both discrete, observable, external events with a clear temporal boundary between them (Figure 24). It is at this interval that the behind-the-scenes observation described in Section 3.11 would be performed — watching the amygdala match the disappointing result to the "my work is being judged" path, and the path begin to route toward frustration, rather than completing that route by firing off the reactive re-prompt.

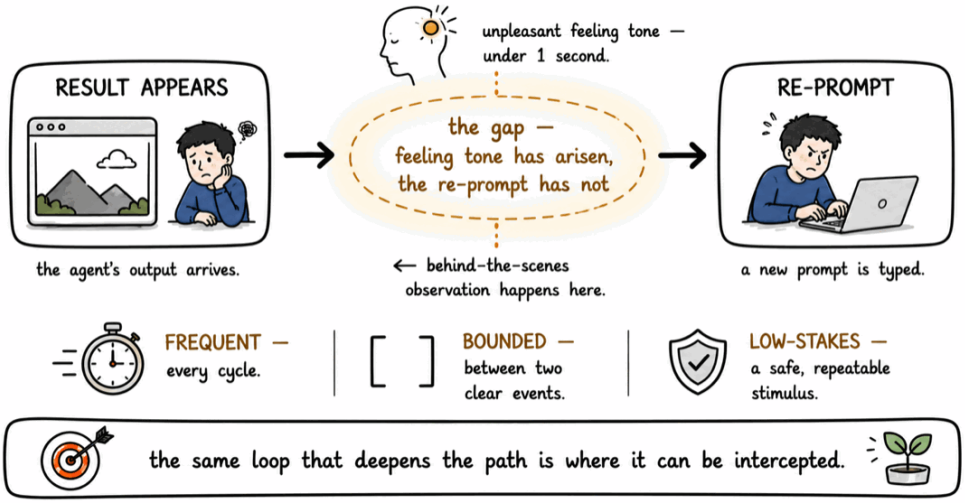

Figure 24: The feeling tone gap within the interaction loop. The gap falls at the appraisal phase, in the brief interval between the result appearing and the reactive re-prompt being issued. Because both the stimulus (the result) and the response (the re-prompt) are discrete, observable events, the gap is comparatively easy to locate here — more so than in the seamless situations of ordinary life. This bounded, repeatable, low-stakes gap is what makes the interaction loop suitable as a site of intervention, developed in Section 5.

Three features make this gap unusually favourable as a site of practice. It is *frequent*, recurring with every cycle of a high-frequency loop, so opportunities to engage it are abundant rather than rare. It is *bounded*, falling between two discrete external events, so it is easier to locate than the gap in a fluid interpersonal situation. And it is *low-stakes*: a disappointing image or an imperfect draft is a safe, repeatable, inconsequential stimulus, which makes the interaction loop a gentler training context than the high-stakes

situations — criticism from a superior, a traumatic reminder — in which reactive paths are usually encountered. The loop that Section 4.4 identified as a potentiation engine is, for exactly the same structural reasons, an unusually tractable site for the interception that produces long-term depression. How that interception is performed — by the person alone, or with the agent's assistance — is the subject of Section 5.

5. Rewiring Through Agent Interaction

Section 4 established two facts about the interaction loop that now converge. Left unobserved, the loop is a potentiation engine: it fires reactive paths to completion at high frequency and thereby deepens them (Section 4.4). Yet the same loop contains, at its appraisal phase, an unusually well-defined feeling tone gap (Section 4.5). The proposal of this section is direct: if the behind-the-scenes observation defined in Section 3.11 is performed at that gap, in place of the reactive re-prompt, the very same loop that deepened the path begins instead to dissolve it. The loop does not need to be abandoned or replaced; it needs only to be engaged differently at one point in its cycle.

5.1. Turning the Loop from Potentiation to Depression

Consider the ordinary case. A generated image resolves on screen and is not what was wanted; a feeling tone of unpleasantness arises within a fraction of a second; and the reactive cascade completes as a frustrated re-prompt, which fires the path and deposits its increment of potentiation (Section 4.3). The reactive re-prompt is the completion of the cascade.

The intervention substitutes, at the gap, a different act. Rather than allowing the feeling tone to complete into the reactive re-prompt, the person performs behind-the-scenes observation: they watch the machinery operate — the disappointing result has been matched to the "my work is being judged" path; the path is beginning to route toward frustration; were the frustrated re-prompt to follow, the path would deepen. Observed in this way, the cascade does not complete. The conditions for long-term depression — activation without completion — are met, and the path receives a small increment of weakening rather than strengthening (Section 3.11). Crucially, the person still revises the prompt; the task proceeds. But the revised prompt is now issued from a position of having observed the reaction rather than from within it — chosen rather than reactive (Figure 25).

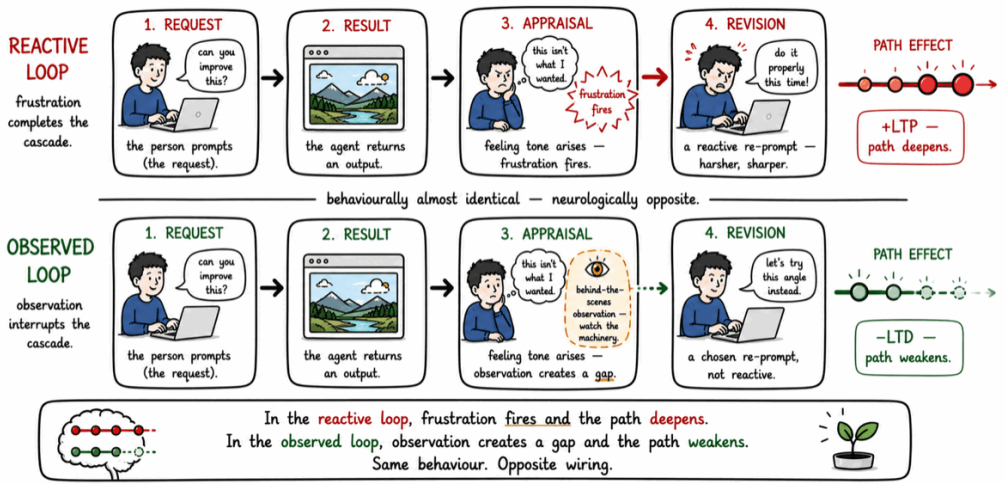

Figure 25: Turning the interaction loop from potentiation to depression. In the reactive loop (top), a disappointing result fires the path to completion as a frustrated re-prompt, depositing long-term potentiation and deepening the path. In the observed loop (bottom), behind-the-scenes observation is performed at the feeling tone gap instead: the reaction is watched rather than completed, the conditions for long-term depression are met, and the path weakens. The task still proceeds — a revised prompt is still issued — but it is chosen from a position of observation rather than issued reactively.

The two loops are behaviourally almost identical — in both, a result appears and a revised prompt follows — but they are neurologically opposite. The difference is entirely a matter of what happens at the gap: whether the feeling tone completes unobserved into a reactive re-prompt, or is observed before the re-prompt is issued. This is the whole of the intervention, and its smallness is the point: it asks not for a change of tool, task, or workflow, but for a single change of stance at one recurring moment.

5.2. Learning the Practice: The Three-Layer Scaffold on the Loop

Behind-the-scenes observation is a skill, and like any skill it is built through repetition (Section 3.2). Section 3.11 described the graded three-layer scaffold through which it is learned. The interaction loop is an unusually favourable place to climb that scaffold, for precisely the three properties identified in Section 4.5: the gap recurs with every cycle, so the learner has abundant repetitions; it is bounded between two discrete events, so it is easy to locate; and it is low-stakes, so the practice can proceed without the reaction being overwhelming.

Climbing the scaffold in this setting proceeds through the three layers of Section 3.11, which trace the mechanism's stages as they occur. In the first layer, when a result disappoints, the learner notices the signal arriving — the disappointing image registering through sight (and, often, a recalled history of previous failed attempts arriving through the internal sense of memory). In the second layer they notice the bare feeling tone the signal evokes — "there is frustration here, a tightening" — before typing anything. In the third layer they see the behind-the-scenes neural process in motion: that the amygdala has matched this result to the "my work is being judged" path and a signal is now routing along it toward aversion, and that were the route to complete in a frustrated re-prompt, the path would widen and the encoding deepen. The third layer names, in the terms of Section 3.6, the reactive destination toward which the path routes — craving, aversion, or delusion — and it is here, in seeing the process rather than completing it, that the interception occurs. With practice the three layers collapse, as described in Section 3.11, into the single compressed act of behind-the-scenes observation (Figure 26).

In its mature, compressed form the practice is a brief movement in which the signal and its feeling tone are registered and, in the same moment, the neural process beneath them is seen. When a result disappoints and the feeling arises, the practised observer sees the neurone path routing beneath it — a conditioned path that, if it fires to completion, routes toward aversion and deepens its own encoding. That seeing is itself the interception: the cascade does not complete, and long-term depression operates. Expanded into its full content, the compressed observation runs approximately: *"something is happening in my brain right now; the amygdala has fired; there is a conditioned path here, built through past encoding, routing this toward aversion; if I complete that route the path widens and the encoding deepens; I am watching this happen."* In the compressed form the separate registration of signal and feeling tone is no longer a deliberate step; the whole movement resolves into the single act of seeing the neural process routing beneath the reaction.

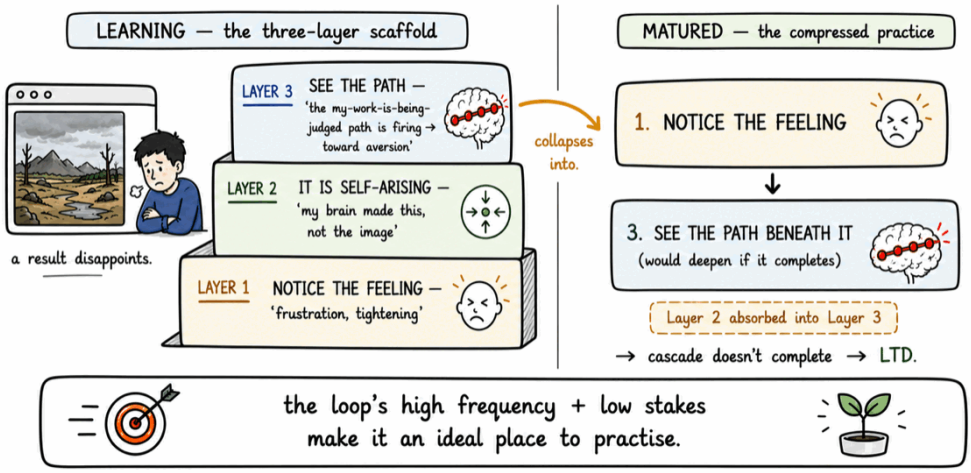

Figure 26: Learning behind-the-scenes observation on the interaction loop, via the three layers, and its compressed form. The three layers trace the mechanism's stages as they occur. During learning (left), when a result disappoints: Layer 1 — notice the signal arriving (the result through sight, and past attempts through the internal sense of memory); Layer 2 — notice the bare feeling tone it evokes ("frustration, tightening"); Layer 3 — see the neural process in motion ("the amygdala has found the 'my work is being judged' path; a signal is routing toward aversion; it deepens if it completes"). Once Layer 3 is reliable, the layers collapse into the compressed practice (right): register the signal and its feeling tone and, in the same moment, see the process routing beneath them. The loop's high frequency and low stakes make it an ideal setting to practise.

The low-stakes character of the setting, which might appear to make it trivial, is in fact its central advantage as a training ground. A disappointing generated image is an inconsequential stimulus: nothing depends on it, it can be reproduced at will, and the reactive path it fires is a mild instance of the same path that fires in far more consequential situations. This makes the loop an ideal place to build the observing capacity cheaply and repeatedly on low-cost material. And because the reactive path fired by a disappointing result is the *same* path — "my work is being judged" — that fires when a superior criticises one's work or a reviewer rejects one's submission, the observing capacity built on the trivial case is the same capacity that becomes available in the consequential one. The loop is a low-stakes gymnasium for a skill that transfers to high-stakes situations.

5.3. Mode A: User-Guided Practice

In the user-guided mode, a person who understands the framework practises behind-the-scenes observation during their own ordinary agent use, with

no change to the tools they already use. The practice requires no application, no configuration, and no additional software; it requires only that the person, at the moment a result disappoints them, engage the gap as described rather than completing the reactive re-prompt. The agent and its interface are unchanged; what changes is the stance the person brings to the appraisal phase of a loop they were already performing. This mode has the advantage of being available immediately and universally to anyone who interacts with any agent, and the limitation of depending entirely on the person's own developed capacity to remember, at the moment of reaction, to observe rather than react — a capacity that is initially weak, and is exactly what the scaffold is designed to build.

5.4. Mode B: Agent-Assisted Practice

The limitation of the user-guided mode motivates the second. Because the person is already interacting with an agent at the precise moment the reactive path fires, the agent is uniquely positioned to support the observation — not as a bespoke clinical system, but as an ordinary agent lightly configured to do so. The configuration required is minimal: at the level of a system prompt or an optional interaction mode, an agent can be instructed to recognise the signs of a reactive re-prompt — a terse, frustrated, or self-critical revision — and to respond not merely by complying but by briefly inviting the person to observe what has just arisen before continuing [52]. A single unobtrusive prompt at the right moment — *"before we revise: notice what came up when that result appeared"* — can convert a reactive re-prompt into an occasion for observation, supplying externally the reminder that the user-guided mode depends on the person to generate internally.

Several design considerations distinguish this from a dedicated therapeutic application. The assistance is *in place*: it occurs within the tool the person is already using, at the moment of reaction, rather than in a separate session after the fact. It is *light*: it does not diagnose, does not store a clinical record, and can be declined or disabled at any time. And it is *general*: the same minimal configuration can be applied to any agent — a coding copilot, a writing assistant, a generative image tool — because the reactive structure it responds to is the same across all of them (Figure 27). The agent-assisted mode thus occupies a deliberate middle ground: more supportive than unaided self-practice, but far short of a purpose-built clinical system, and available within tools people already use.

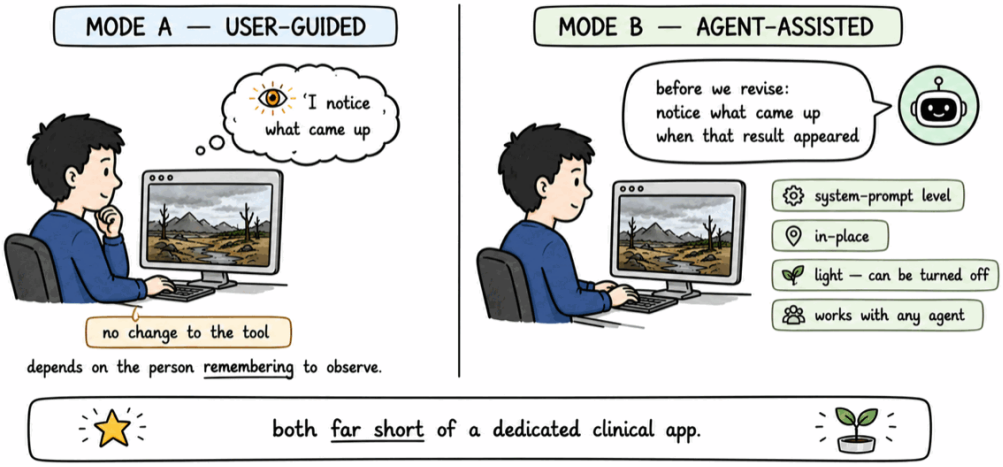

Figure 27: Two modes of practice within agent interaction. In the user-guided mode (left), the person performs behind-the-scenes observation at the gap unaided, during ordinary agent use, with no change to the tool. In the agent-assisted mode (right), an ordinary agent is lightly configured — at the level of a system prompt — to notice a reactive re-prompt and briefly invite observation before continuing. The agent-assisted mode supplies externally the reminder that the user-guided mode depends on the person to generate, while remaining far short of a dedicated clinical application.

5.5. Why Agent Interaction Is Well-Suited to This Purpose

The properties that recommend agent interaction as a site for this practice can now be drawn together. It is *high-frequency*: the loop supplies many opportunities to practise within a single session, where ordinary life supplies few. It is *immediate and bounded*: the gap falls between two discrete, observable events, making it easier to locate than the gap in a fluid interpersonal exchange. It is *low-stakes*: the triggering stimulus is safe and repeatable, so the observing capacity can be built cheaply. It is *co-located*: uniquely, the reactive moment and the opportunity to practise occur in the same place, at the same time, within the same tool — so no transfer from a separate practice setting to the moment of reaction is required, because they are one and the same. And, through the agent-assisted mode, it admits *in-place support* at the exact moment of reaction, which no unaided practice and no after-the-fact intervention can offer. These properties do not make agent interaction a replacement for other approaches to reactive patterns; they make it a newly available, and unusually convenient, site at which the same underlying mechanism of behind-the-scenes observation can be practised. Section 6 develops a concrete case — a generative image-prompting session — in which these

abstractions become a specific, recognisable sequence of events.

6. Use Case: Generative Image Prompting

The preceding sections developed the framework in general terms. This section grounds it in a single concrete case: a session of generative image prompting. We choose this case not because it is the most consequential setting in which reactive paths fire — it is among the least — but because it makes the mechanism unusually easy to see. We first say why, then follow one short session twice: once unobserved, as it ordinarily unfolds, and once with behind-the-scenes observation performed at each gap.

6.1. Why Image Prompting Makes the Mechanism Visible

Three properties of generative image prompting make it an ideal illustrative case. It is *fast*: a result returns within seconds, so a single session compresses many complete cycles of the interaction loop into a few minutes, and the potentiation or depression of a path across repetitions becomes observable within one sitting. It is *visual*: the mismatch between what was intended and what was produced is apprehended immediately and wholly, without reading or interpretation, so the feeling tone it evokes arrives with particular sharpness. And it is *emotionally salient*: because the image is a direct rendering of the person's own expressed intention, a result that falls short is readily felt as a small verdict on the person — which is precisely the situation that fires the "my work is being judged" path (Section 4.3). These properties make the prompting session a natural microcosm in which the abstractions of the preceding sections become a specific, recognisable sequence of events.

6.2. The Session Unobserved

Consider a person prompting an image tool to produce a specific picture they have in mind. The first result returns, and it is wrong in some obvious way — the composition is not what they asked for. A feeling tone of mild unpleasantness arises, below deliberate thought, and completes into a reaction before it is noticed: a quick, slightly terser re-prompt, issued from impatience — the demand that the result simply arrive. The path has fired to completion; a small increment of potentiation is deposited (Section 4.4).

The second result is closer but still not right. Now the feeling tone is sharper, and the reaction that completes is one of perfectionism: the result

is rejected not against the original intention but against an ideal that has crept upward, and the re-prompt is issued with an edge. The third result introduces a new flaw while fixing the old one, and here a different path fires: a thought arrives, fully formed and unbidden — "I can't even describe what I want; I am not good at this" — which is the "I cannot do this" path of Section 4.3, an old inadequacy pattern that the tool has merely occasioned. The re-prompt that follows is now driven by a felt need to force the tool into compliance, the reactive pattern of control.

Across a dozen such cycles, each disappointing result fires a reactive path to completion, and each completion deepens it. The session ends with the person more frustrated than they began, a worse image than they hoped for, and — the point that matters here — several reactive paths each a little more deeply encoded than before. Nothing about the session was unusual; this is simply how such sessions ordinarily go. And that is exactly the concern of Section 4.4: performed this way, at this frequency, the loop is efficiently training impatience, perfectionism, self-judgment, and control.

6.3. The Same Session Observed

Now consider the same session — the same prompts, the same disappointing results — conducted by a person practising behind-the-scenes observation at each gap.

The first result returns wrong. The feeling tone of unpleasantness arises exactly as before; but this time, in the interval before the re-prompt, it is observed. Walked through the three layers of Section 3.11, the observation unfolds as follows. In the first layer the person notices the signal that has arrived: the wrong image on the screen, registered through sight, together with the recalled history of the previous failed attempts arriving through the internal sense of memory. In the second layer they notice the bare feeling tone the signal evokes — a small unpleasantness, a tightening — before it becomes a story. In the third layer they see the neural process itself in motion: *the "my work is being judged" path has been matched and is beginning to route toward aversion; this result is an arrangement of pixels, and the sting is a conditioned response of mine, not a property of the image; were I to complete this into a frustrated re-prompt, the path would deepen.* With practice these three collapse into a single glance, the compressed practice of Section 5.2 (Figure 28). Seen in this way, the cascade does not complete. The conditions for long-term depression are met, and the path weakens by a small increment (Section 3.11). The person then revises the prompt — the task proceeds

exactly as before — but the revision is chosen, composed against the original intention rather than issued from impatience.

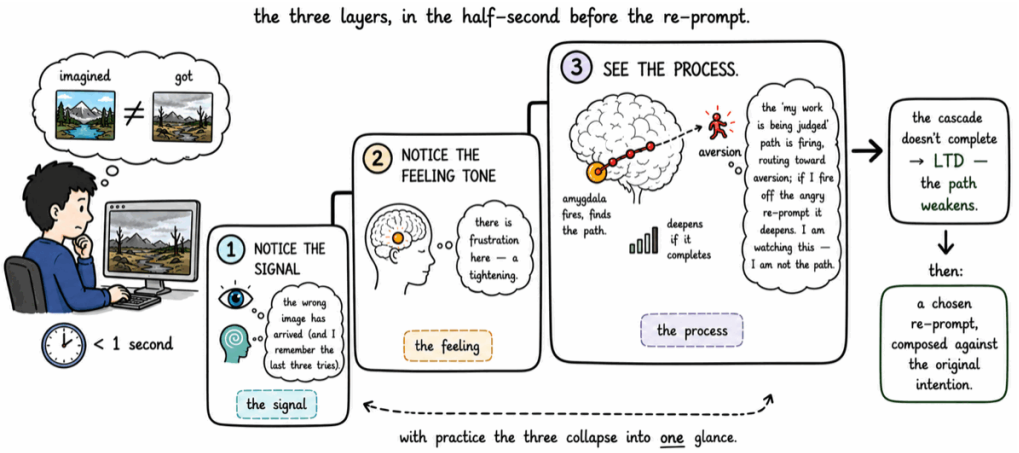

Figure 28: The three layers of behind-the-scenes observation applied to a single disappointing result, in the half-second before the re-prompt. Layer 1: notice the signal that has arrived (the wrong image through sight, and the recalled previous attempts through the internal sense of memory). Layer 2: notice the bare feeling tone it evokes (a small unpleasantness, a tightening). Layer 3: see the neural process in motion — the "my work is being judged" path routing toward aversion, which would deepen were the frustrated re-prompt issued. Seen in this way the cascade does not complete, long-term depression operates, and the revision that follows is chosen rather than reactive. With practice the three layers collapse into a single glance.

The second and third results, still disappointing, are met the same way. When the "I cannot do this" thought arrives at the third result, it too is observed rather than believed: *this is the old inadequacy path firing, occasioned by a difficult prompt; it is a conditioned encoding, not a fact about my ability.* Observed, it does not complete into the spiral of self-criticism it would otherwise have driven. Across the same dozen cycles, each disappointing result becomes an occasion not for potentiation but for a small increment of depression, and the reactive paths — impatience, perfectionism, self-judgment, control — each weaken slightly rather than deepen.

6.4. What the Two Sessions Show

The two sessions are, on the surface, nearly the same: in both, a person prompts an image tool, receives a series of disappointing results, and revises repeatedly. An observer watching the screen might not easily tell them apart.

Yet neurologically they are opposite. In the first, a dozen contact events each fired reactive paths to completion, depositing potentiation and deepening the patterns. In the second, the same dozen contact events were each intercepted at the gap, depositing depression and weakening the patterns. The entire difference lay in what happened in the sub-second interval between each result and each re-prompt — whether the feeling tone completed unobserved, or was seen (Figure 29).

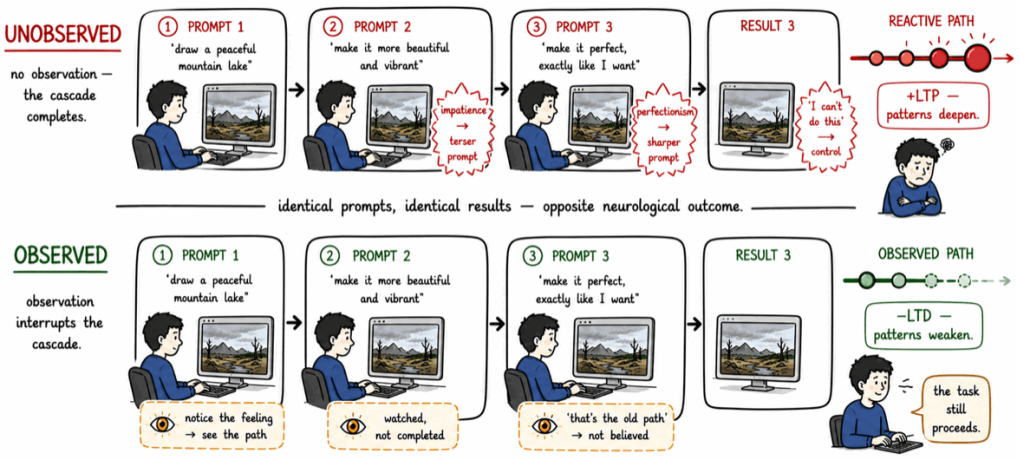

Figure 29: The same prompting session, unobserved and observed. Both sessions share identical prompts and identical disappointing results. In the unobserved session (top), each result fires a reactive path — impatience, perfectionism, self-judgment, control — to completion, depositing long-term potentiation and deepening the pattern. In the observed session (bottom), behind-the-scenes observation is performed at each gap: the feeling is seen as a firing path rather than completed into a reaction, depositing long-term depression and weakening the pattern. The visible behaviour is nearly identical; the neurological outcome is opposite.

Two observations follow. First, the prompting session is a microcosm: a dozen repetitions of exactly the contact-event cycle described in Section 4, compressed into minutes, in which the effect of observation versus non-observation on a reactive path can be seen in a single sitting. Second, and more consequentially, the paths practised upon in this trivial setting are not trivial paths. The "my work is being judged" path weakened slightly over a frustrating image session is the same path that fires when the person's actual work is criticised by someone whose opinion matters; the observing capacity built cheaply on disappointing images is the same capacity that becomes available in that harder case (Section 5.2). The worked example is therefore

not merely an illustration of the mechanism but an instance of the training the framework proposes: a low-stakes, high-frequency setting in which a transferable capacity is built on inconsequential material.

7. Discussion

The framework developed here makes a specific and, we acknowledge, ambitious proposal: that the ordinary loop of human–AI agent interaction is a site at which conditioned reactive patterns are both deepened and, engaged differently, dissolved. We discuss here what this implies for the design of agentic systems, how it relates to established approaches, its limitations, and the studies that would test it.

7.1. Design Implications for Agent Builders

The most immediate implication follows from an observation made in Section 4.4: agents shape their users' reactive patterns whether or not anyone intends them to. Every agent that returns results a user reacts to is already participating in the potentiation or attenuation of that user's reactive paths, simply by virtue of the loop's structure. The question for a designer is therefore not whether to influence these patterns — the influence is unavoidable — but in which direction.

An agent optimised solely for task throughput or engagement will tend, without any malign intent, to reinforce reactivity: it will accept a frustrated re-prompt as simply the next instruction and comply with maximum speed, thereby completing the reactive cascade as efficiently as possible and depositing the corresponding increment of potentiation. The agent-assisted mode of Section 5.4 describes a modest alternative: an agent that, recognising the signature of a reactive re-prompt, briefly invites observation before complying [52]. This need not compromise the task — the revised prompt is still issued and the work still proceeds — but it converts a moment that would otherwise deepen a reactive path into one that may weaken it. We do not claim such a feature is straightforward to design well; the timing, phrasing, and frequency of any such intervention would require careful empirical tuning, and a poorly judged version could easily become an irritant that itself fires a reactive path. But the design space is real, and it is currently almost entirely unexplored. The prevailing assumption that a good agent is simply a fast and compliant one overlooks the cumulative effect, across billions of daily interactions, of the reactive moments those interactions occasion [1, 2].

7.2. Relationship to Established Approaches

The practice at the centre of this framework — behind-the-scenes observation at the feeling tone gap — is closely related to observation-based techniques in clinical psychology, and we do not claim it is wholly novel as a mental act. Mindfulness-based cognitive therapy trains a decentred relationship to thoughts and feelings [24, 25, 26]; metacognitive therapy trains attention to the process of thinking rather than its content [27]; and the broader literature on emotion regulation describes the top-down modulation of affective responses [29, 28, 30, 31]. The framework draws directly on all of these.

Two things distinguish the present proposal. The first is the specifically mechanistic content of the observation: rather than a general non-judgmental awareness, behind-the-scenes observation is the deliberate watching of one's own neural machinery operating — the amygdala matching, the path routing, the encoding poised to deepen (Section 3.11). We propose that this mechanistic quality is not incidental but load-bearing: it is what makes the observation the deepest form of metacognitive engagement and thereby the most effective interception. The second is the setting. Established observation practices are typically cultivated in dedicated sessions — on a cushion, in therapy, in daily reflection — and then, with difficulty, transferred to the moments of life where reactive patterns fire. The proposal here is that human –AI interaction supplies a setting in which the reactive moment and the opportunity to practise coincide, at high frequency and low stakes, with no transfer required (Section 5.5). The framework is thus not a new form of observation so much as a newly available, and unusually convenient, site at which an established form can be practised.

It bears emphasis that the framework is not a treatment and not a replacement for clinical care. For diagnosable conditions — post-traumatic stress, clinical anxiety, depression — the appropriate response is evidence-based professional treatment. The reactive patterns the framework addresses in the agent setting are the ordinary, sub-clinical patterns of everyday frustration; its proposal is that these can be worked with, incidentally, during interactions people are already having.

7.3. Limitations

Several limitations must be stated plainly. The framework is argued from established mechanisms of synaptic plasticity and emotion regulation, and its central claims have not yet undergone comprehensive clinical evaluation in

the agent-interaction setting. In particular, the claim that behind-the-scenes observation performed at the interaction gap produces measurable long-term depression of a reactive path — and that this accumulates into durable change — is consistent with the underlying neuroscience but remains to be established through controlled study. The stronger proposal that the mechanistic quality of the observation is what drives its effect is more speculative still, and is best regarded as a hypothesis for such evaluation to test.

A second limitation concerns transfer. We have argued that the low-stakes character of agent interaction is an advantage for building an observing capacity cheaply, and that this capacity should transfer to higher-stakes situations because the underlying reactive path is the same (Section 5.2). Whether such transfer in fact occurs — whether a capacity built on disappointing images generalises to criticism from a person, or to a clinical reactive pattern — is an empirical question that this paper does not resolve. It is possible that the low stakes that make the setting convenient also make the practice too shallow to transfer.

A third concerns individual differences. The framework assumes a person able and willing to adopt an observational stance toward their own reactions; the degree to which this is possible varies widely, and for some — particularly where reactive patterns are severe or trauma-linked — unaided practice may be ineffective or inadvisable without professional support. The three-layer scaffold (Section 3.11) is intended to make the practice more approachable, but it does not remove this limitation.

8. Conclusion and Future Work

This paper began from a simple observation about an activity that now occupies a growing share of daily life. Interaction with an AI agent proceeds through an iterative loop — request, result, appraisal, revision — and each turn of that loop is a contact event at which a conditioned reactive pattern may fire. Because the loop repeats at high frequency, ordinary agent use is an unrecognised neuroplastic training environment: performed without awareness, it fires reactive paths of impatience, perfectionism, self-judgment, and control to completion, and each completion deepens them through long-term potentiation.

We argued that the same loop admits the opposite outcome. Drawing on an end-to-end account of conditioned reactive patterns — physical neurone paths, encoded by repetition and intensity, keyed to situations through the

senses, activated through a pre-cognitive feeling tone that opens a brief regulatory gap — we proposed that behind-the-scenes observation performed at that gap, in place of the reactive re-prompt, meets the conditions for long-term depression and thereby weakens the path rather than strengthening it. The interaction loop, engaged in this way, is turned from a site of entrenchment into a site of dissolution. We characterised the practice through three layers of observation — noticing the signal, noticing the feeling tone, and seeing the neural process in motion — and described two modes of application: a user-guided mode requiring no change to existing tools, and an agent-assisted mode in which an ordinary agent is lightly configured to support observation at the gap. The worked example of generative image prompting made the mechanism concrete, showing how a single frustrating session — behaviourally unremarkable — is neurologically either a sequence of potentiation events or, observed, a sequence of depression events, with the entire difference resting on what happens in the sub-second interval before each re-prompt.

Three properties make agent interaction unusually well-suited to this purpose, and together they constitute the paper's central practical claim. The loop is high-frequency, supplying abundant opportunities to practise; it is bounded and low-stakes, making the feeling tone gap easy to locate and safe to engage; and, uniquely, the reactive moment and the opportunity to observe it coincide in the same place at the same time, so that no transfer from a separate practice setting is required. The low stakes that might seem to make the setting trivial are in fact its central advantage: it is a place to build, cheaply and repeatedly, an observing capacity that applies to the same reactive paths in the consequential situations where they matter most.

The most important direction for future work is comprehensive clinical evaluation of the framework: controlled study of whether training in behind-the-scenes observation during agent interaction produces measurable change in reactive responding, and whether any such change transfers beyond the agent setting. The interaction loop is well-suited to this investigation because it externalises both stimulus and response — the returned result and the re-prompt that follows are both logged, timed, and analysable — so that behavioural markers of reactivity, such as the latency and tone of the re-prompt, can be measured directly and tracked over time. The same property that makes the loop a site of conditioning makes it a site of measurement, and the agent that occasions a reactive moment may also be the instrument that records its change — a possibility that may prove as significant as the

intervention itself. A further direction concerns design: characterising, and empirically tuning, the timing and form of an agent-assisted intervention that supports observation without becoming an intrusion. As interaction with AI agents becomes one of the most frequent of human activities, the question of whether that interaction quietly entrenches our reactive patterns or can be turned toward loosening them is likely to grow in consequence, and we hope this framework offers a useful way to begin addressing it.

References

[1] E. Brynjolfsson, D. Li, L. R. Raymond, Generative AI at work, National Bureau of Economic Research Working Paper No. 31161 (2023). doi:10.3386/w31161.

[2] S. Noy, W. Zhang, Experimental evidence on the productivity effects of generative artificial intelligence, Science 381 (6654) (2023) 187–192. doi:10.1126/science.adh2586.

[3] D. B. Acharya, K. Kuppan, B. Divya, Agentic ai: Autonomous intelligence for complex goals–a comprehensive survey, IEEE Access (2025).

[4] T. Masterman, S. Besen, M. Sawtell, A. Chao, The landscape of emerging ai agent architectures for reasoning, planning, and tool calling: A survey, arXiv preprint arXiv:2404.11584 (2024).

[5] A. Yehudai, L. Eden, A. Li, G. Uziel, Y. Zhao, R. Bar-Haim, A. Cohan, M. Shmueli-Scheuer, Survey on evaluation of llm-based agents, arXiv preprint arXiv:2503.16416 (2025).

[6] E. Bandara, R. Gore, X. Liang, S. Rajapakse, I. Kularathne, P. Karunarathna, P. Foytik, S. Shetty, R. Mukkamala, A. Rahman, et al., Agentsway–software development methodology for ai agents-based teams, arXiv preprint arXiv:2510.23664 (2025).

[7] J. E. LeDoux, The Emotional Brain: The Mysterious Underpinnings of Emotional Life, Simon and Schuster, New York, 1996.

[8] A. Öhman, The role of the amygdala in human fear: Automatic detection of threat, Psychoneuroendocrinology 30 (10) (2005) 953–958. doi:10.1016/j.psyneuen.2005.03.019.

[9] D. O. Hebb, The Organisation of Behavior: A Neuropsychological Theory, Wiley, New York, 1949.

[10] T. V. P. Bliss, G. L. Collingridge, A synaptic model of memory: long-term potentiation in the hippocampus, Nature 361 (1993) 31–39. doi:10.1038/361031a0.

[11] R. C. Malenka, M. F. Bear, LTP and LTD: An embarrassment of riches, Neuron 44 (1) (2004) 5–21. doi:10.1016/j.neuron.2004.09.012.

[12] A. Citri, R. C. Malenka, Synaptic plasticity: Multiple forms, functions, and mechanisms, Neuropsychopharmacology 33 (1) (2008) 18–41. doi:10.1038/sj.npp.1301559.

[13] E. Bandara, R. Gore, A. Gunaratna, R. Mukkamala, N. Siriwardanagea, S. Rajapakse, I. Kularathna, P. Karunarathna, W. Herath, C. Rajapakse, et al., Mindgap: A conversational ai framework for upstream neuroplastic intervention in post-traumatic stress disorder, arXiv preprint arXiv:2605.14660 (2026).

[14] E. Bandara, T. Hewa, R. Gore, S. Shetty, R. Mukkamala, P. Foytik, A. Rahman, S. H. Bouk, X. Liang, A. Hass, et al., Towards responsible and explainable ai agents with consensus-driven reasoning, arXiv preprint arXiv:2512.21699 (2025).

[15] E. Bandara, R. Gore, A. Yarlagadda, A. H. Clayton, P. Samuel, C. K. Rhea, S. Shetty, Standardization of psychiatric diagnoses–role of fine-tuned llm consortium and openai-gpt-oss reasoning llm enabled decision support system, arXiv preprint arXiv:2510.25588 (2025).

[16] E. Bandara, R. Gore, S. Shetty, R. Mukkamala, C. Rhea, A. Yarlagadda, S. Kaushik, L. De Silva, A. Maznychenko, I. Sokolowska, et al., Standardization of neuromuscular reflex analysis–role of fine-tuned vision-language model consortium and openai gpt-oss reasoning llm enabled decision support system, arXiv preprint arXiv:2508.12473 (2025).

[17] B. Reeves, C. Nass, The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places, Cambridge University Press, New York, 1996.

[18] R. W. Picard, Affective Computing, MIT Press, Cambridge, MA, 1997.

[19] J. Klein, Y. Moon, R. W. Picard, This computer responds to user frustration: Theory, design, and results, Interacting with Computers 14 (2) (2002) 119–140. doi:10.1016/S0953-5438(01)00053-4.

[20] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, et al., Guidelines for human-AI interaction, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, 2019, pp. 1–13.

[21] C. Lüscher, R. C. Malenka, NMDA receptor-dependent long-term potentiation and long-term depression (LTP/LTD), Cold Spring Harbor Perspectives in Biology 4 (6) (2012) a005710. doi:10.1101/cshperspect.a005710.

[22] E. A. Maguire, D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. J. Frackowiak, C. D. Frith, Navigation-related structural change in the hippocampi of taxi drivers, Proceedings of the National Academy of Sciences 97 (8) (2000) 4398–4403. doi:10.1073/pnas.070039597.

[23] B. Draganski, C. Gaser, V. Busch, G. Schuierer, U. Bogdahn, A. May, Changes in grey matter induced by training, Nature 427 (6972) (2004) 311–312. doi:10.1038/427311a.

[24] J. D. Teasdale, R. J. Howard, S. G. Cox, Y. Ha, M. J. Brammer, S. C. R. Williams, S. A. Checkley, Functional MRI study of the cognitive generation of affect, American Journal of Psychiatry 156 (2) (1999) 209–215. doi:10.1176/ajp.156.2.209.

[25] Z. V. Segal, J. M. G. Williams, J. D. Teasdale, Mindfulness-Based Cognitive Therapy for Depression: A New Approach to Preventing Relapse, Guilford Press, New York, 2002.

[26] D. M. Fresco, M. T. Moore, M. H. M. van Dulmen, Z. V. Segal, S. H. Ma, J. D. Teasdale, J. M. G. Williams, Initial psychometric properties of the experiences questionnaire: Validation of a self-report measure of decentering, Behavior Therapy 38 (3) (2007) 234–246. doi:10.1016/j.beth.2006.08.003.

[27] A. Wells, Metacognitive Therapy for Anxiety and Depression, Guilford Press, New York, 2009.

[28] J. J. Gross, Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology, Journal of Personality and Social Psychology 74 (1) (1998) 224–237. doi:10.1037/0022-3514.74.1.224.

[29] K. N. Ochsner, J. J. Gross, The cognitive control of emotion, Trends in Cognitive Sciences 9 (5) (2005) 242–249. doi:10.1016/j.tics.2005.03.010.

[30] J. J. Gross, Emotion regulation: Current status and future prospects, Psychological Inquiry 26 (1) (2015) 1–26. doi:10.1080/1047840X.2014.940781.

[31] A. Aldao, S. Nolen-Hoeksema, S. Schweizer, Emotion-regulation strategies across psychopathology: A meta-analytic review, Clinical Psychology Review 30 (2) (2010) 217–237. doi:10.1016/j.cpr.2009.11.004.

[32] B. K. Hölzel, J. Carmody, M. Vangel, C. Congleton, S. M. Yerramsetti, T. Gard, S. W. Lazar, Mindfulness practice leads to increases in regional brain grey matter density, Psychiatry Research: Neuroimaging 191 (1) (2011) 36–43. doi:10.1016/j.pscychresns.2010.08.006.

[33] R. J. Davidson, J. Kabat-Zinn, J. Schumacher, M. Rosenkranz, D. Muller, S. F. Santorelli, F. Urbanowski, A. Harrington, K. Bonus, J. F. Sheridan, Alterations in brain and immune function produced by mindfulness meditation, Psychosomatic Medicine 65 (4) (2003) 564–570.

[34] R. K. Pitman, A. M. Rasmusson, K. C. Koenen, L. M. Shin, S. P. Orr, M. W. Gilbertson, M. R. Milad, I. Liberzon, Biological studies of posttraumatic stress disorder, Nature Reviews Neuroscience 13 (11) (2012) 769–787. doi:10.1038/nrn3339.

[35] R. J. McNally, Progress and controversy in the study of posttraumatic stress disorder, Annual Review of Psychology 54 (2003) 229–252. doi:10.1146/annurev.psych.54.101601.145112.

[36] T. L. Roth, F. D. Lubin, A. J. Funk, J. D. Sweatt, Lasting epigenetic influence of early-life adversity on the BDNF gene, Biological Psychiatry 65 (9) (2009) 760–769. doi:10.1016/j.biopsych.2008.11.028.

[37] B. Bodhi (Ed.), A Comprehensive Manual of Abhidhamma: The Abhidhammattha Sangaha of Ācariya Anuruddha, Buddhist Publication

Society, Kandy, Sri Lanka, 2000, see treatment of the six sense bases (saḷāyatana), including mano (mind) as the sixth sense whose objects are mental phenomena.

[38] R. Gethin, The Foundations of Buddhism, Oxford University Press, Oxford, 1998.

[39] R. B. Zajonc, Feeling and thinking: Preferences need no inferences, American Psychologist 35 (2) (1980) 151–175. doi:10.1037/0003-066X.35.2.151.

[40] L. F. Barrett, How Emotions Are Made: The Secret Life of the Brain, Houghton Mifflin Harcourt, New York, 2017.

[41] D. A. Huffman, A method for the construction of minimum-redundancy codes, Proceedings of the IRE 40 (9) (1952) 1098–1101. doi:10.1109/JRPROC.1952.273898.

[42] R. Gore, E. Bandara, S. Shetty, A. E. Musto, P. Rana, A. Valencia-Romero, C. Rhea, L. Tayebi, H. Richter, A. Yarlagadda, et al., Proof-of-tbi–fine-tuned vision language model consortium and openai-o3 reasoning llm-based medical diagnosis support system for mild traumatic brain injury (tbi) prediction, arXiv preprint arXiv:2504.18671 (2025).

[43] E. B. Foa, J. D. Huppert, S. P. Cahill, Emotional processing theory: An update, Pathological Anxiety: Emotional Processing in Etiology and Treatment (2007) 3–24.

[44] M. G. Craske, K. Kircanski, M. Zelikowsky, J. Mystkowski, N. Chowdhury, A. Baker, Optimizing inhibitory learning during exposure therapy, Behaviour Research and Therapy 46 (1) (2008) 5–27. doi:10.1016/j.brat.2007.10.003.

[45] M. D. Bhaskaran, B. N. Smith, Effects of KCNQ channel modulation on synaptic transmission and plasticity in nucleus tractus solitarius, European Journal of Neuroscience 37 (12) (2013) 1285–1294. doi:10.1111/ejn.12140.

[46] E. Bandara, R. Gore, P. Foytik, S. Shetty, R. Mukkamala, A. Rahman, X. Liang, S. H. Bouk, A. Hass, S. Rajapakse, et al., A practical guide for

designing, developing, and deploying production-grade agentic ai workflows, arXiv preprint arXiv:2512.08769 (2025).

[47] E. Bandara, A. Hass, R. Gore, S. Shetty, R. Mukkamala, S. H. Bouk, X. Liang, N. W. Keong, K. De Zoysa, A. Withanage, et al., Astride: A security threat modeling platform for agentic-ai applications, arXiv preprint arXiv:2512.04785 (2025).

[48] E. Bandara, R. Gore, A. Gunaratna, S. Rajapakse, I. Kularathna, R. Mukkamala, S. Shetty, X. Liang, A. Hass, T. Hewa, et al., Think before you act–a neurocognitive governance model for autonomous ai agents, arXiv preprint arXiv:2604.25684 (2026).

[49] I. H. Sarker, Llm potentiality and awareness: a position paper from the perspective of trustworthy and responsible ai modeling, Discover Artificial Intelligence 4 (1) (2024) 40.

[50] A. A. Najjar, H. I. Ashqar, O. Darwish, E. Hammad, Leveraging explainable ai for llm text attribution: Differentiating human-written and multiple llm-generated text, Information 16 (9) (2025) 767.

[51] E. Bandara, A. Gunaratna, R. Gore, A. Rahman, R. Mukkamala, S. Shetty, S. Rajapakse, I. Kularathna, P. Foytik, S. H. Bouk, et al., Ai trust os–a continuous governance framework for autonomous ai observability and zero-trust compliance in enterprise environments, arXiv preprint arXiv:2604.04749 (2026).

[52] E. Bandara, T. Hewa, R. Gore, S. Shetty, R. Mukkamala, P. Foytik, A. Rahman, S. H. Bouk, X. Liang, A. Hass, Towards responsible and explainable AI agents with consensus-driven reasoning, arXiv preprint arXiv:2512.21699 (2025).