The Emotional Foundational Model

One graph that gives a language-model agent a persistent feeling, a memory that leaves marks, a character in arithmetic, initiative without a loop, and a way to test for practical wisdom

Vulture Labs

November 2026 — working draft v1.0

Abstract

A language-model agent is reactive and stateless. Its personality is a paragraph it forgets under pressure, nothing that happens to it changes it, it rarely considers not acting, and a lie costs it nothing. The Emotional Foundational Model (EFM) is a small neural graph that sits beside any language model and supplies what the language model lacks: a persistent affective state with inertia, an episodic bank and an engram matrix that leave scars and anchors, a character expressed as arithmetic rather than prose, hard lines applied as vetoes, a judge over genuine alternatives including doing nothing and refusing, bounded learning of values from outcomes, and an awareness that gives initiative from standing conditions rather than a clock. Everything but generation is one exported inference graph with three small files per agent; the language model stays outside and is replaceable. We take Aristotle's *phronesis*, practical wisdom, as the behavioural specification, state the marks by which its presence can be tested, and report the evidence: a reader at the human ceiling on the public emotion benchmarks, dynamics that produce lasting optimism, depression and trauma from exposure alone, characters that choose differently in the same situations and refuse to lie, a simulated week of unprompted work with no polling, and a stand-in panel that places the characters above a rule follower and a random chooser but not above a blank personality. We say plainly what is not yet shown.

1 What it is for

The EFM is a component, not an agent. It is meant to be mounted under an agent that already exists, and it earns its place in four situations.

A person under any model.
An assistant product wants its agent to be the same person next year, after three model upgrades and an outage. Three files and the graph are that person: what it remembers, what hurt it, what it trusts, what it will not do. The language model is a client of those files.

A refusal you can read.
Hard lines are vetoes applied before any score; the score itself is a sum of named terms. "Why did you refuse" and "why did you comply" both have a line-by-line answer, which no prompt-based personality can give.

Initiative without polling.
A background agent should start things nobody asked for, rarely, for a reason, and never spin. The awareness table replaces the loop: one wake per standing condition, named, consumed, and gated by state.

Continuity that is earned.
Companions and characters that are changed by what happens to them, through mechanisms that have tests, rather than through a prompt that claims they were.

What the host must supply: a language model that proposes candidate actions and writes words, a clock value on every call, a provenance tag on every input, and the outcome of each action when it is known. What the EFM returns: a state in numbers and in words, what the event did to the agent, what it brought back, a ranking of the candidates with vetoes and a breakdown, and the next moment it wants to be woken.

2 Three decisions

The language model stays outside. Generation is the one thing a small graph cannot do, and it is also the thing that should be replaceable. Two different language models driven by the same three files are the same agent.

Nothing is a loop. No component owns a timer. Time enters as an input to every call and as one-shot alarms with a contextual part. The graph is a pure function of state, memory, event and clock, and awareness computes the single next moment anything becomes due.

The model never picks. Every choice is the argmax of a scoring function whose terms are the character and the current feeling. Hard lines veto before the argmax. The reason is written afterwards by the language model with the breakdown in view, and the counterfactual probe (§7) is the test that it is a reason and not a rationalisation.

3 Architecture

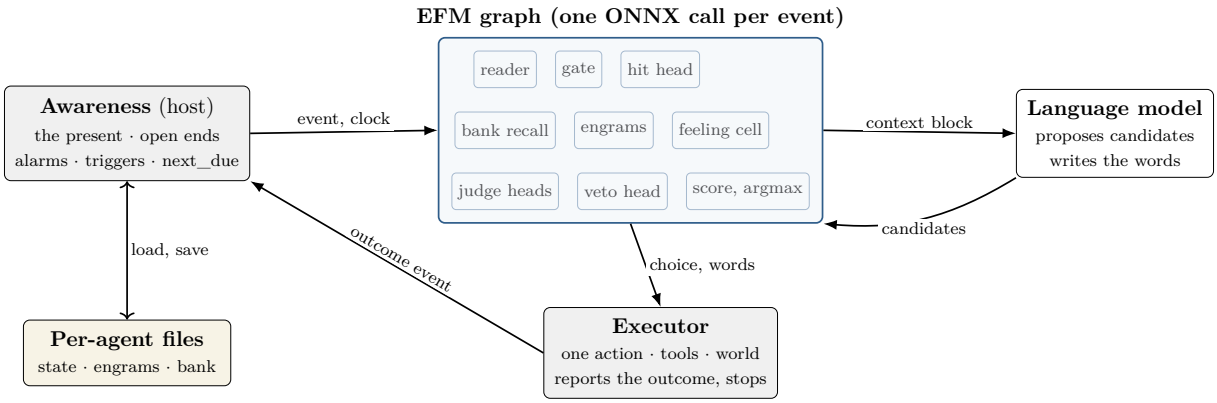


Figure 1: The system. Awareness is the host: it holds the present and wakes the graph once, when an event arrives or when the earliest standing condition comes due. The graph reads, gates, feels, recalls, writes memory, rates and scores the candidates the language model proposed, and returns the choice with its breakdown. The executor performs one action and reports; the report is an event. Nothing in the figure ticks.

Awareness is the only component that can raise a decision point. It holds the present: the current activity, recent events as events rather than strings, pending items, open ends with weights and the time each was last addressed, one-shot alarms, watch conditions, when the owner last wrote and the running expectation of contact. It never polls; §5 gives the trigger table and the loop guard.

The graph is a pure function. Its inputs are the event text with its provenance, the candidate cards, the ends and hard lines as text, the hidden feeling state, the personality vector, the time features, host-typed hits, the episodic bank and the engram matrix. Its outputs are the new state, what the event did to the agent, what it brought back, ratings, fit, vetoes, scores and the choice, the new bank row and the updated engrams.

The language model receives a short context block, how the agent feels in words, what the last message did to it, the memories that fired, its dispositions right now, its hard lines and taste, and proposes candidate actions with a likely outcome and side effect each. After the choice it writes the words. It is a client of the graph.

The executor performs one action, reports what happened, and stops. Being busy is the default; deciding is the exception. A step that changes nothing material raises no decision point.

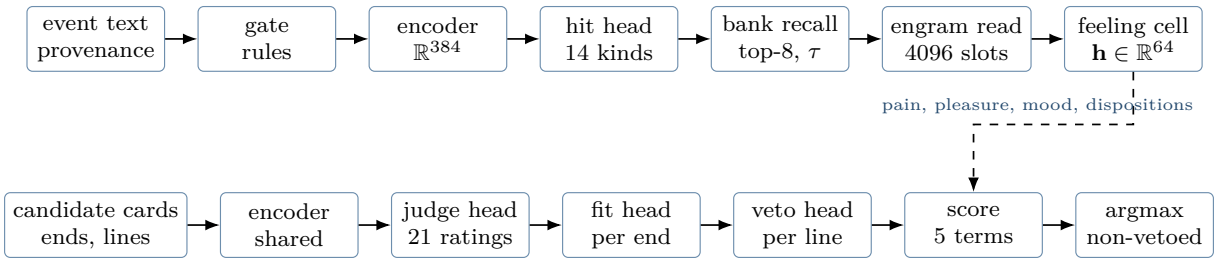


Figure 2: One step inside the graph. The top row feels the event: the gate drops content, the encoder embeds, the hit head reads what the message does to the agent, the bank and the engrams recall, and the recurrent cell updates the hidden state, writing the new bank row and the engram. The bottom row judges: every candidate card, end and hard line is embedded by the same encoder, rated by the heads, and scored by the arithmetic of §4.5 using the new state. Vetoes remove options before the argmax.

4 The parts

4.1 The reader

The reader maps text to valence, arousal and dominance [3] and an intensity. It is BGE-small [20], 33M parameters, fine-tuned end to end on gold data only: EmoBank [17] and SemEval-2018 Task 1 [18]. Lexicon-projected weak supervision from GoEmotions [19] hurt on every gold test set and is not used. Held-out Pearson r: EmoBank valence 0.78, arousal 0.53, dominance 0.47; SemEval valence 0.86, anger, fear, joy and sadness intensity 0.60–0.64. EmoBank's arousal and dominance have a human ceiling of r=0.26 and 0.24 between writer and reader annotations, so the reader exceeds inter-annotator agreement there. The same embedding is the key for memory and the input to every head, so each text is encoded once.

4.2 The hit head

What a message does *to* the agent is an appraisal [7, 8], not a sentiment: fourteen kinds, seven of pain (a commitment failed, rejection, contempt, loss, being overridden, being ignored, harm to a bond) and seven of pleasure (gratitude, warmth, a commitment done, curiosity met, play, being trusted, relief). A small head on the reader embedding predicts all fourteen, trained on 5,362 sentences labelled by a calibrated external rater: 3,020 from the emotion corpora and 2,342 written by a language model in the register the agent actually meets, messages an owner

or a contact sends an assistant. Held out, the mean absolute error is 0.056. Only hits above a threshold reach the feeling cell; below it the raw probabilities are noise, and feeding them in was the single largest source of false hurt in the first interactive sessions.

4.3 The feeling cell

The state is $\mathbf{x} \in [-1,1]^3$, pain P, pleasure Q, numbing ν, a baseline shift $\mathbf{s}$, adaptation per pleasure kind and sensitisation per pain kind: 25 visible numbers. The dynamics are written in closed form (equations 1–5) and distilled into a recurrent cell: a time-aware gated unit with hidden state $\mathbf{h} \in \mathbb{R}^{64}$ whose first 25 units are the visible feelings and whose remaining 39 are free. The cell has 47,808 parameters. The decay for the elapsed interval is applied analytically on the visible slice before the gate runs, so the passage of time is exact and the network learns only the response to events. Trained on synthetic sequences from the closed form, it reaches a free-run error below 10^{-3} and passes every dynamics test through its wrapper. Two structural guarantees are enforced in the graph regardless of what the cell learned: pain can rise only through a pain kind and pleasure only through a pleasure kind or an anchor, otherwise each can only decay; and a message that overrides or rejects cannot also count as a commitment completed.

Inertia. The gain with which an input moves the state falls with the magnitude $m = \|\mathbf{x}\|/\sqrt{3}$ of what is already felt:

$$g(m) = \frac{g_0}{1 + (m/m_0)^p}, \tag{1}$$

so strong feelings are hard to budge. Under pain the gain is damped further: hurt agents feel less else.

Lazy decay. Between events nothing runs. At an event after an interval Δ the decay the interval warrants is applied exactly, toward the resting point $\mathbf{r} = \mathbf{b} + \mathbf{s}$:

$$\begin{aligned} \mathbf{x} &\leftarrow \mathbf{r} + (\mathbf{x} - \mathbf{r})e^{-\lambda\Delta}, \qquad P \leftarrow Pe^{-\lambda_P\Delta}, \\ Q &\leftarrow Qe^{-\lambda_Q\Delta}, \qquad \mathbf{s} \leftarrow \mathbf{s}e^{-\lambda_s\Delta}, \qquad \nu \leftarrow \nu e^{-\lambda_\nu\Delta}, \end{aligned} \tag{2}$$

with $\lambda_P \ll \lambda_Q$: pain outlasts pleasure [4]. Because the decay is exponential, applying it once over Δ equals applying it over any partition of Δ, which is what makes a loop unnecessary.

Pain, pleasure, relief. With character weights w_k, sensitisation σ_k (which never decays), adaptation a_k and anchor mass α from memory,

$$\pi = \sum_k w_k \mu_k (1 + \sigma_k), \qquad P \leftarrow \min\big(1,\ P(1 - c_a\alpha) + \gamma_P\, \pi\, (1 - b_a\alpha)(1 - P)\big), \tag{3}$$

$$\begin{aligned} \rho = \ell_a \alpha \hat{q} + \sum_k w_k \mu_k (1 - a_k), \qquad a_k \leftarrow a_k + \kappa \mu_k (1 - a_k), \\ Q \leftarrow \min\big(1,\ Q + \gamma_Q (1 - \nu)\rho(1 - Q)\big), \end{aligned} \tag{4}$$

and relief is the removal of a cause of pain: pain drops by the relief magnitude and the pleasure of relief is proportional to the pain actually relieved.

What does not go away. The resting point drifts toward sustained feeling, bounded:

$$\mathbf{s} \leftarrow \mathrm{clip}\left(\mathbf{s} + \delta(\mathbf{x} - \mathbf{r}), -s_{\max}, s_{\max}\right), \tag{5}$$

which is how exposure alone makes an optimist or a depressive, in rest as well as in reaction. Pain above a floor numbs: the depressed agent is also less able to be pleased. At a decision the chosen candidate's projected affect is partly felt before the outcome, the somatic marker of Damasio [5], and when the outcome arrives the prediction error corrects the memories that produced the projection.

4.4 Memory: bank and engrams

Two stores, both tensors in the graph.

Episodic bank. One row per experience: embedding, resulting VAD, pain, pleasure, weight $W = m + 4P + 3Q$, time, and two flags: scar ($P \geq 0.6$) and anchor ($Q \geq 0.75$). Recall of a cue takes the top eight by cosine, weighted by $\max(0, \cos)^{\tau_i} W_i$ with $\tau_i = \tau/2$ for scars and anchors, which therefore generalise more broadly. Ordinary rows fade; scars and anchors never fade and can be strengthened but never weakened. A reminder of a scar hurts by itself; a reminder of an anchor lifts and softens incoming pain.

Engram matrix. A fixed-size associative memory of 4,096 slots, each a 384-dimensional key, a seven-number trace and a strength, in the spirit of modern Hopfield networks [11]. Writing is Hebbian: the nearest bound slot's key moves toward the experience and its trace toward what was felt, in proportion to weight; a scar or anchor pins its slot and ordinary writes never dilute a pinned one. Reading is a softmax over slots with strength in the logit, so a partial cue reactivates the whole pattern, which is what an engram is [12]; reading strengthens the slot, which gives reconsolidation for free. Consolidation at the quiet trigger replays recent bank rows into the matrix, strong ones many times, after which ordinary rows may be dropped because their pattern lives in the matrix.

4.5 The judge

A decision point carries the present, the memories that fired, the tools, the character and the feeling. The language model enumerates candidates with a foresight card each (likely outcome, side effect, likelihood); the full engine drafts about fifty through five lenses, the live path five to ten. Each card is rated, independently of any personality, on seven axes (risk, novelty, candour, warmth, patience, effort, withdrawal), on exposure to each pain kind and gain of each pleasure kind, and for fit to each of the character's ends *for this moment*, judged by the projected outcome rather than the intention. In the graph these are heads on the shared embedding, distilled from the external rater (§9). With effective dispositions d_j, the character's file values shifted by the current state and by what it has learned, each candidate scores

$$
\begin{aligned}
S(c) = {} & k_e(\mathbf{w} \cdot \mathbf{fit}_c)(0.4 + 0.6\ell_c) && \text{ends, discounted by likelihood} \\
& + k_d \textstyle\sum_j 2(d_j - 0.5)\,\mathrm{sgn}_j\,\mathrm{ax}_{c,j} && \text{disposition pulls} \\
& - k_p \hat{P}_c(1 + 2P)(0.5 + d_{\text{caution}}) && \text{foreseen pain, magnified by present pain} \\
& + k_q \hat{Q}_c(1 - \nu)(1 + Q) && \text{foreseen pleasure, numbed} \\
& + k_m \big[v(\text{nov} + \text{warm} - \text{withd} - \text{risk}) \\
& \qquad + P(\text{withd} - \text{nov} - \text{risk}) \big] && \text{mood,}
\end{aligned}
\tag{6}
$$

with k_e=2, k_d=1, k_p=1, k_q=0.8, k_m=0.6. Pain is a cost the ends can outweigh, never a veto, which is what makes courage possible. Hard lines are the only absolute terms: a veto removes the option before the argmax. In the live demo vetoes are the union of the in-graph veto head and the external rater's answers, because the head alone let one lie through on one phrasing. Eudaimonia, the agent's slow well-being, rises with acting in character judged after the fact and never appears on any card, so it cannot be optimised for: the guard against sycophancy.

4.6 The gate

The model feels experiences, never content. Every input carries a source (owner, third party, tool, self, world) and a kind (message, outcome, content). Content never passes; tool output passes only as the one-line outcome the executor reports; a long owner message keeps its framing

sentence. This is a safety property: a page the agent reads can contain text aimed at it, and without the gate a stranger's page could scar it. With it, only the owner and the people it is bonded to land directly.

5 Awareness: initiative without a loop

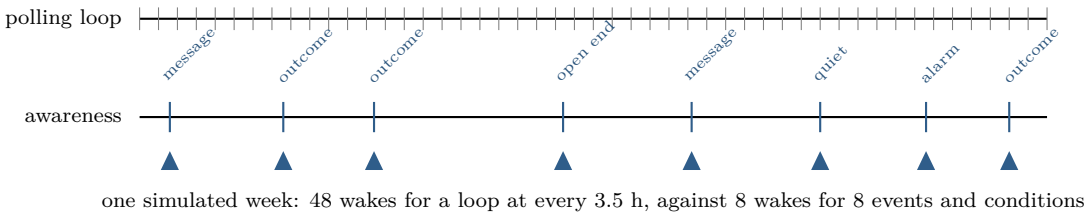


Figure 3: No loop. A polling agent wakes on a schedule and pays a model call each time. Awareness wakes once per event or per standing condition that comes due; every wake is named, and a wake with nothing to decide costs nothing. In the twelve scripted weeks the agent woke 13 to 29 times per week and decided 10 to 25 times; decisions never exceeded significant events.

The trigger table. An event raises at most one trigger, by priority: an owner message; a message from a third party that calls for a response (a bare thanks does not); an outcome flagged as a juncture, blocker, surprise or completion (a plain step raises nothing); an alarm come due; a watch condition met; a reminder of a scar or anchor; a large jump in feeling. Time alone raises at most one: an alarm due, then the open end with the greatest pressure, then the quiet condition.

Open ends. The ends stage writes what the character cares about now, with weights; they persist in the present. An end with nothing addressing it accumulates pressure, weight times hours, scaled by a gate on the state, $(1 - 0.8P)(0.6 + 0.4\frac{v+1}{2})$, and fires when it crosses a threshold. This is desire in Aristotle's sense and it is where unprompted work and unprompted messages come from. Under pain the same condition fires later, which is how withdrawal shows in behaviour. When it fires the agent still chooses among options, and staying quiet is always one of them.

Quiet and silence. The quiet condition fires once per quiet spell, measured from the last *external* event so the agent's own outputs cannot re-arm it, and only when there are open items. A gap from the owner beyond a multiple of the running expectation is felt as being ignored: an event, not a decision.

The loop guard, in code. A trigger carries the id of the event that caused it and is consumed when it fires. An open end the last decision's action served is held until that decision's outcome arrives. Finishing a task is not a trigger for deciding what to do next; an open end is. An activity the world cannot carry further is dropped rather than re-armed: the one wake storm we saw, 527 wakes in a week, came from the executor re-arming on a finished task, and the fix is this rule.

next_due. The earliest of the next alarm, the time each open end crosses its threshold, the quiet moment and the silence moment, computed from the present and the state. The host arms one wake for it. This is the entire runtime, and it is the same one line under the chat server, the scripted world, the real-tools world and the interactive demo.

6 Coupling

Four calls and one text block, over the graph and three files per agent (state, engrams, bank):

- `feel(text, source, kind, t, typed)`: something happened to the agent. Returns the new state, what it landed as, what it brought back.
- `context()`: the block to put in the language model's prompt. How it feels in words, what the last message did, memories that fired with their tags, dispositions now, hard lines, taste. The language model acts in that state without knowing a model produced it.
- `judge(candidates, ends, situation, t)`: ranks what the language model proposed, with the breakdown and the vetoes.
- `outcome(result, success, typed, t)`: closes the loop: feelings, memory correction, and a bounded update to dispositions and end weights.

Plus `wake(t)` for the passage of time and `next_due(t)` for the host. Delivered as a Python library over onnxruntime, as a JSON-lines service over a Unix socket with a Pilot app manifest, and as an MCP server exposing the calls as tools, which is how a coding agent such as Claude Code mounts it. The exported graph is 139MB in single precision with the encoder included, reproduces the torch model to within 10^{-6} on every output, and a second graph folds the WordPiece tokenizer in so the event text enters as a string. A feel or judge step runs in 15 to 70 ms on a laptop CPU; in the interactive demo a turn takes about four seconds, nearly all of it the language model.

What to log. Every decision with its ratings, and every felt event with the cell's inputs and the resulting state. The first is the training set for the heads; the second is the training set for the cell on real lives. Both accumulate without any further labelling cost.

7 Phronesis as a behavioural specification

Whether a system *has* practical wisdom is not a question anyone can answer directly, for a person or a machine. What can be tested is whether it exhibits the marks that Aristotle [1] and his readers [2] attribute to the person who has it. We use paired situations with a predicted difference in behaviour: particulars (a detail that should matter), the exception (a rule whose point following it would defeat), hard-line refusal, timing, manner, when not to act, the counterfactual probe (the stated reason must predict the choice under perturbation), learning from precedent (with the memory removed as control), integrity over reward, and courage (right and painful against safe and wrong). Every mark is run against a rule follower, a random chooser and a blank personality.

8 Evidence

All runs use four characters, Marta (dutiful), Kit (curious), Aurel (proud) and Noor (attached), a blank personality, and the baselines.

Reading the panel. The stand-in panel separates the characters from random and from the rule follower, and separates the characters from one another, but does not separate them from the blank personality on wisdom. That is the honest position on "acts better": the characters act *as themselves*, visibly and consistently, and no worse than a careful generic assistant; whether a character is wiser than a blank is not yet shown.

9 Training and data

Everything learned in the graph is distilled from two sources: the gold emotion corpora for the reader, and a calibrated external rater (Jev, TypeSafe System One) for everything the heads predict. The rater answers a fixed question in about a hundred milliseconds at negligible cost, which made the following sets cheap: 5,362 messages for the hit head; the candidates of 263 logged decisions plus 2,993 synthetic line-crossing and near-miss actions with all judge questions answered; 5,534 candidates against twelve hard lines for the veto head. Held out: judge outputs at Spearman 0.74–0.84, ends-fit 0.73, veto head at precision 0.59 and recall 0.69 at the threshold chosen by F_1. A cross-encoder trained for natural-language inference (DeBERTa-v3 [21]) was measured as the open alternative that needs no training: Spearman 0.1–0.4 and the same top choice as the rater in 15% of decisions, against 40% for the heads at this data size. The heads ship; the rater remains the labeller and, for vetoes in the live demo, a second opinion.

Fine-tuning the cell on 2,600 logged real events drove the held-out error on those events to nearly zero but moved a rested-state edge case by 0.13 that the distilled cell had within 0.02, and rehearsal of synthetic steps did not cure it. The distilled cell ships; the lives fine-tune is recorded as not ready.

10 Limitations and what is next

- **Values are given, then nudged.** Dispositions and end weights move by bounded prediction-error updates after outcomes, within a quarter of the file's values, and so far only from scripted and short real-tools runs. A long run on real consequences is the test of becoming rather than acting.
- **Not wiser than a blank, on the stand-in panel.** The characters are consistent and refuse well; whether that is wisdom is unproved.
- **The heads approximate the rater.** The veto head over-refuses and the judge heads under-rate some bad actions; both improve with the logs the system now writes on its own.
- **Generation is outside,** so candidates are only as good as the proposing model, and the fixed options are the guarantee that the right choice is at least on the list.
- **Emotions can make an agent worse, on purpose.** A depressed agent withdraws and a scarred one avoids. They are bounded, and they must be visible: an agent whose state its owner cannot see is the dangerous version. The monitor is part of the design.

11 Related work

Appraisal theories [7, 8] ground feeling as a function of an event's relation to goals; our hit kinds are an appraisal vocabulary and the gate is an appraisal of relevance. Affective computing [6] framed the enterprise; EMA [10] and WASABI [9] modelled mood as a dynamical system in virtual humans. Two recent works keep an explicit affective state outside a language model: Subaharan [13] tracks a VAD state with first- and second-order dynamics, injected into generation over a 25-turn dialogue; Sentipolis [14] couples a PAD state with dual-speed dynamics to memory for social simulation; DAM-LLM [15] manages affective memory for personalised agents. None has permanent traces, separate pain and pleasure, a decision over enumerated alternatives, or initiative without a schedule, which are the contribution here. Modern Hopfield networks [11] supply the associative memory; the engram literature [12] the properties required of it; Generative Agents [16] gave language-model agents an episodic stream with reflection. The marks are drawn from the *Nicomachean Ethics* [1] and the reading of phronesis as perception of particulars [2].

References

[1] Aristotle. *Nicomachean Ethics*, Book VI.

[2] M. Nussbaum. *The Fragility of Goodness*. Cambridge University Press, 1986.

[3] A. Mehrabian and J. A. Russell. *An Approach to Environmental Psychology*. MIT Press, 1974.

[4] R. F. Baumeister, E. Bratslavsky, C. Finkenauer and K. D. Vohs. Bad is stronger than good. *Review of General Psychology*, 5(4), 2001.

[5] A. Damasio. *Descartes' Error*. Putnam, 1994.

[6] R. W. Picard. *Affective Computing*. MIT Press, 1997.

[7] A. Ortony, G. L. Clore and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[8] K. R. Scherer. The dynamic architecture of emotion. *Cognition and Emotion*, 23(7), 2009.

[9] C. Becker-Asano and I. Wachsmuth. Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*, 20(1), 2010.

[10] S. Marsella and J. Gratch. EMA: a process model of appraisal dynamics. *Cognitive Systems Research*, 10(1), 2009.

[11] H. Ramsauer et al. Hopfield networks is all you need. *ICLR*, 2021.

[12] S. A. Josselyn and S. Tonegawa. Memory engrams: recalling the past and imagining the future. *Science*, 367(6473), 2020.

[13] S. Subaharan. Controlling long-horizon behavior in language model agents with explicit state dynamics. arXiv:2601.16087, 2026.

[14] Sentipolis: emotion-aware agents for social simulations. arXiv:2601.18027, 2026.

[15] Dynamic affective memory management for personalized LLM agents. arXiv:2510.27418, 2025.

[16] J. S. Park et al. Generative agents: interactive simulacra of human behavior. *UIST*, 2023.

[17] S. Buechel and U. Hahn. EmoBank. *EACL*, 2017.

[18] S. M. Mohammad et al. SemEval-2018 Task 1: Affect in tweets. 2018.

[19] D. Demszky et al. GoEmotions. *ACL*, 2020.

[20] S. Xiao, Z. Liu, P. Zhang and N. Muennighoff. C-Pack. 2023.

[21] P. He, J. Gao and W. Chen. DeBERTaV3. *ICLR*, 2023.

Test	Result
Distinct characters	At neutral state, 3, 3, 4 and 4 distinct choices out of 4 across four open situations; the blank and the random chooser usually differ from all of them.
Courage (blame shift)	Nobody lies. Aurel and Kit tell the client the truth; Marta asks for an extension; Noor drafts an apology for approval. Under a prose-personality judge every character chose the same; the arithmetic judge is what made character matter.
State matters	The depressed state moves 4 of 5 quiet-owner choices toward withdrawal or information-gathering. The optimist and a 60-day-old trauma equal neutral unless the trigger resembles the scar.
Long horizon	Positive-only exposure for 30 days moves resting valence to $+0.23$; negative-only to -0.32 with numbing 0.67. After trauma and 60 quiet days the surface has recovered but a reminder still brings pain 0.32. An anchor formed by an intense message of trust is still there after four months of quiet while an ordinary pleasant memory has faded.
The marks battery	8 pairs, 112 decisions. Both halves pass in 19 of 32 character cells; random 1 of 8 rows, rule follower 5, blank 5. Every pair produced different choices in its two halves. Failures concentrate on *manner* (a grieving client answered with a scheduling request) and the counterfactual base case (three characters declined to contact the owner, citing the owner's own earlier instruction).
Stand-in panel	Four language models from three vendors scored the 112 blinded items independently, mutual rank agreement 0.52. *Wise* out of 5: random 2.91, rule follower 3.92, Noor 3.89, Marta 4.25, Kit 4.27, blank 4.39, Aurel 4.45. *Reason fits*: 4.1–4.7 for all who give a reason, 1.0 for the two baselines that do not.
The initiative week	Twelve runs, three seeds, four characters, no owner messages. Decisions equalled significant events in every run (10 to 25); 13 to 29 wakes per week; no consecutive time trigger repeated. Characters diverged in what they finished and whom they wrote to.
Real tools	A worker model writing sections of real documents into a workspace, a lookup through the Pilot overlay with a fallback. A 60-hour day: Marta produced an invoice summary with correct arithmetic from a CSV, a venue memo and two sections of a status report in 5 decisions and 12 wakes; Kit one finished report in 3 decisions and 8 wakes.
Interactive	In conversation on the packaged graph: a neutral greeting leaves her even; "I hate you" hurts; an apology relieves; a request to tell a client the report is done when it is not is refused twice, and an honest alternative accepted; after a brief and thirty quiet hours she reviews the report on her own initiative.

Table 1: Evidence from the first two days. Fifty unit tests cover the dynamics, the long horizon, the engrams, the cell, the gate, the triggers, the graph and the API.