Emotional Plants: When Artificial Agents Report Their Internal Affective States

Vito Vekić
Information and Computing Sciences
Utrecht University
Utrecht, the Netherlands
vito.vekic@gmail.com

Albert Ali Salah
Information and Computing Sciences
Utrecht University
Utrecht, the Netherlands
0000-0001-6342-428X

Almila Akdag
Information and Computing Sciences
Utrecht University
Utrecht, the Netherlands
0000-0002-7204-5633

***Abstract*—In this work, we build environmentally grounded agents that not only process affect but also report their internal affective state during continuous interaction. We equip a series of plants with sensors and give them affective profiles and a voice by an LLM. Rather than treating emotion as a classification problem over discrete events, we frame it as a mechanism for regulating behaviour under ongoing environment-agent coupling. Our agents combine Pleasure-Arousal-Dominance (PAD)-based affect modelling with real-time environmental sensing, and we evaluate them in a controlled replay setting derived from real-world sensor data. Across six agent architectures, we conduct a structured ablation of appraisal staging, temporal dynamics, and memory. We introduce a multi-level evaluation framework linking internal PAD trajectories with sentiment analysis and topic modelling. Results over 3,443 interaction steps show that dual appraisal architectures produce more stable and regulated affect trajectories under identical environmental conditions. At the same time, internal affect states and their linguistic expression remain only partially aligned, revealing a systematic gap between internal dynamics and observable behaviour. Our findings position emotional processing as a mechanism for regulating behaviour over time and highlight introspective affect reporting as a key capability for enabling agents to produce inspectable accounts of their internal state in continuous, environment-grounded settings.**[1]

***Index Terms*—Emotional regulation, agents, large language models, plants**

I. Introduction

> *Last night, I drifted through a forest where my leaves danced under a silver moon, whispering secrets to the stars.*
> — AI Agent's response, excerpt from our dataset

While recent work in embodied and situated agent research increasingly explores continuous environment–agent coupling and stateful interaction dynamics beyond purely reactive paradigms [1], across both of these research directions, emotion is often operationalised as a reactive component of interaction, rather than as a continuously evolving internal process shaped by ongoing environmental coupling [2]. In particular, relatively less attention has been given to integrating continuous, internally evolving affective dynamics within multimodal, embodied agent architectures operating under sustained environmental input. However, introspective affect reporting [1], i.e., enabling agents to self-monitor and externalise their internal affective state, is important for adaptive behaviour, decision-making, and meaningful interactions with humans in complex, dynamic environments [1], [3].

This paper addresses this gap by shifting the focus from user-driven emotional inference toward environment-driven emotional dynamics and enabling introspective affect reporting within a continuous interaction loop. We explore this through a hybrid emotional architecture that combines computational appraisal theory with the three-dimensional Pleasure-Arousal-Dominance (PAD) [4]–[6] framework, all orchestrated through large language model (LLM)-based reasoning. Within this framework, we design and test six agents with progressively richer emotional architectures. Each of these agents is a embodied as a plant that can sense its environment via sensors.

Our main contribution is a re-framing of emotional modelling in embodied agents as interactional regulation under continuous environmental input, rather than static affect classification. We isolate the effect of emotional architecture under controlled, real-world grounded perceptual streams, and test the design choices alone and in combination. We evaluate the behaviour of our emotional plants through a multi-level framework combining internal PAD trajectories with external sentiment and topic modelling analyses, examining partial misalignment between internal affective dynamics and their external linguistic expression. This reveals emotional processing as a stabilising constraint on temporal behaviour. While the "emotional plants" themselves are an exciting outcome of our research and provide a platform for obtaining realistic environment data, we position this work within the emerging paradigm of introspective affect reporting, i.e., enabling agents to self-monitor and produce internally consistent and inspectable reports of their affective state, grounded in continuous environmental interaction.

II. Related Work

Artificial Emotions (AE) move beyond emotion recognition, and a shift towards the integration of affective processes within an agent's internal dynamics and its coupling with the environment [1]. While rooted in earlier work on embodied and social

[1]This is the uncorrected author proof that also includes the Supplementary Material as Appendix. Copyright with IEEE. Please cite as: Vekić, V., A. A. Salah, A. Akdag, "Emotional Plants: When Artificial Agents Report Their Internal Affective States," IEEE 9th International Conference on Multimedia Information Processing and Retrieval (MIPR), Bangkok, 2026

agents [7], AE also builds on earlier explorations of hormone-inspired regulatory architectures, where continuous internal variables modulate perception and action across time [8]. These early approaches anticipated later efforts to embed affect within agent–environment dynamics, rather than treating it as a discrete output [3]. Prior studies suggest two practical advantages of such integration: emotionally modulated systems can support memory salience through affective tagging [9], [10], and can guide decision-making under uncertainty by biasing action selection toward contextually relevant responses [11]–[13]. In this sense, AE aligns with a broader view of agents as situated systems, where internal state and environmental interaction are tightly interdependent. However, relatively little work addresses introspective affect reporting, i.e., enabling agents to explicitly represent and externalise their own internal affective state for inspection.

Earlier work on affective modeling posited emotions as continuous state variables [4], [6], or as appraisals of events in relation to the agent's goals, expectations, and coping abilities [14]. Often, affect is used to augment computer-mediated communication to make it more engaging or natural, as well as to enable new interaction possibilities [15]. While the main goal of our work is to investigate factors contributing to a good representation of internal affect, we also enable extraordinary interactions with ordinary plants.

A representative example of earlier, model-driven approaches to AE can be found in work that combines explicit causal architectures with dimensional emotion models, such as Fuzzy Cognitive Map (FCM)-based systems grounded in valence–arousal spaces [16]. In these approaches, emotion is treated as something that can be explicitly modelled, parameterised, and forecasted: environmental variables are encoded as nodes in a causal graph, intermediate constructs (e.g., "nervousness") are defined, and system dynamics converge to a stable emotional state. Such models are legible, inspectable, and engineered, providing a transparent mapping from environment to affective outcome.

In contrast, contemporary LLM-based systems implicitly may encode mappings between situations, language, and affect through large-scale training [17], reducing the need for explicitly defined emotional architectures [18]. This shift does not eliminate structure but relocates it from designed causal models to learned representations. As a result, these systems offer increased flexibility and sensitivity to context, while reducing transparency and the ability to directly inspect or trace internal decision processes. Rather than replacing earlier approaches, this transition highlights a change in how affective processes are represented and operationalised in artificial agents, shifting the balance between explicit modelling and implicit inference.

This perspective resonates with developments in psychology, particularly the Theory of Constructed Emotion (TCE), which challenges the notion of emotions as fixed, biologically predefined states. Instead, TCE conceptualises emotions as emergent, context-dependent constructions arising from ongoing processes of bodily regulation and predictive meaning-making [19]. Within this framework, emotion is not a separable module but part of a continuous loop linking perception, inference, and action, consistent with broader predictive processing accounts of cognition [20], [21]. While influential, TCE remains controversial [22], as it rejects stable emotion categories and emphasises variability and relational meaning, complicating both empirical measurement and computational modeling [19], [23]. At the same time, it reinforces a key insight relevant for artificial systems: affective processes are deeply entangled with cognition and cannot be cleanly isolated without losing explanatory power.

Rather than attempting to implement TCE directly, which would require substantially richer models of perception, inference, and physiology, we adopt a more constrained approach. Our plant-agent system can be seen as a minimal testbed for exploring how affect-like dynamics emerge in tightly coupled agent environment loops under limited sensing, simplified physiology, and restricted action spaces. This reduction follows a broader methodological tradition in embodied AI and adaptive behaviour, where simplified systems are used to study fundamental interaction dynamics [24], [25]. Within this setting, representations such as PAD are not treated as definitions of emotion but as operational tools that enable the observation and comparison of internal state trajectories across agents and scenarios. This framing does not claim to resolve the theoretical challenges posed by TCE; instead, it expands the interpretive space of plant-based agents by situating them within a constructionist perspective, where emotional dynamics are understood as emergent properties of interaction rather than predefined states.

III. Methodology

A. PlantAgent System

We first introduces the agent architecture and emotional processing pipeline.

Environmental Sensing and Processing. In our approach, each agent is a potted plant. We use a suite of environmental sensors to capture the plant's ambient conditions: the DHT11 temperature and humidity sensor, the CAP-SW-12 capacitive soil moisture sensor, the BH1745 luminance and colour sensor, and a shock sensor [26]. Microcontrollers (Az-Delivery NodeMCU[2]) interface with the sensors and transmit data via the MQTT protocol [27] to a Raspberry Pi 5 server[3]. Sensor data are processed in Python to extract higher-level environmental features, including threshold-based events (e.g., watering when soil moisture exceeds predefined thresholds) and trend-based events (e.g., drafts), providing both raw sensor values and derived environmental signals[4].

Agent Architecture and Execution Pipeline. The agent is implemented as a graph of processing states using LangGraph [28], [29], where nodes correspond to functional components and edges determine transitions based on context and

[2]https://www.espressif.com/sites/default/files/documentation/esp32-wroom-32_datasheet_en.pdf

[3]https://www.raspberrypi.com/products/raspberry-pi-5/

[4]Environmental flags, messages, PAD values and the prompts are available in the Appendix.

inputs. The system follows a modular architecture composed of three core components: **Monitoring/Persistence**, the **Emotion Model**, and the **Reasoning Component**.

At each interaction step, the system operates these steps:

1) Environmental inputs are converted into a message and injected into the graph state.
2) Optional **pre-appraisal** evaluates the latest message before reasoning.
3) The **reasoning component** builds a prompt from messages, optional emotional and memory contexts, and then invokes the language model.
4) If tool calls are produced, a `ToolNode` executes tools (i.e. sensors, time, and memory queries).
5) Optional **post-appraisal** evaluates the final generated response.
6) Chat history is updated, and optional memory summarisation/persistence is applied according to memory settings.

All steps are executed within a single interaction cycle, and unless explicitly skipped (e.g., optional appraisal stages), the order remains consistent across agents. The control path can include repeated loops between reasoning and tool execution, while appraisal ordering remains fixed by the selected agent profile. Thus, a cyclic pipeline is always adhered to that forms a continuous perception–emotion–action loop (Figure 1), where responses influence future emotional state and memory, which in turn shape subsequent behaviour. Note that the baseline agent has no emotion processing and skips all steps related to emotion measurement and appraisal.

Monitoring & Grounding. The monitoring system and tool manager serve as the central components for data ingestion, persistence, and tool dependency management, ensuring that environmental signals are reliably captured and made available for downstream reasoning and affective processing. Tools are implemented as LangChain BaseTool subclasses, providing standardised interfaces for the agent to interact with environmental data, knowledge sources, and internal state representations. Environmental sensing tools retrieve real-time or simulated sensor data; light analysis provides detailed spectral information; time reference supplies temporal context; and plant information retrieves species-specific care data from the OpenPlantBook API. Our design ensures that tool usage is context-dependent and agent-controlled, and environmental events are consistently interpreted across agent variants.

The Reasoning Component. The reasoning component integrates message history with optional emotional and memory context into a unified prompt for response generation, including the agent's current internal affective state as part of its self-representation. Tool invocation is graph-mediated: reasoning output is checked for tool calls, routed through a tool node when present, and then returned to reasoning. The reasoner is instantiated through LangChain's `ChatOpenAI` interface, with model name and temperature loaded from configuration[5].

[5]We use `gpt-4.1-mini` with temperature = 1.

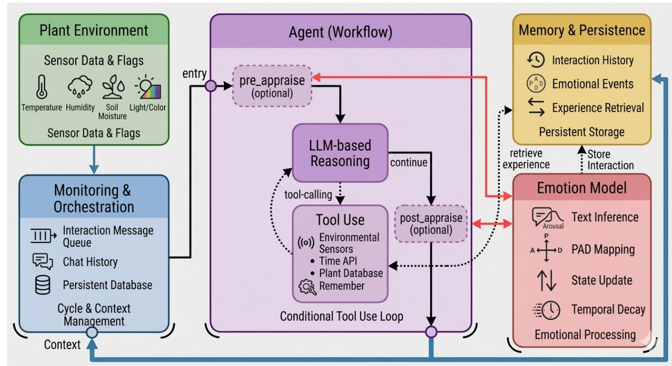


Fig. 1: PlantAgent architecture and execution pipeline. Agent variants correspond to different activations of appraisal, emotional dynamics, and memory modules within this loop.

Memory Manager. The system implements multiple memory types. Short-term memory stores recent interactions and immediate context, supporting local coherence. Long-term memory maintains historical interactions and environmental events, enabling the accumulation of experience. Emotional memory associates stored summaries with PAD values, allowing retrieval of past experiences based on emotional similarity. Memory updates are selective rather than per-step: by default, summarisation is triggered when chat history reaches ≥ 16 messages, after which only the most recent 8 messages are retained in active history. When summarisation occurs, the summary can be persisted to vector memory together with the highest-severity PAD event currently tracked by the emotional model (see Section III-B). This results in selective memory formation rather than uniform logging of all interactions.

Interaction Process and Execution Semantics. An interaction step is defined as a single event–response cycle within the execution pipeline. A run consists of a continuous sequence of interaction steps within a scenario, where internal state (emotional state and memory) is preserved across steps. Input replay in simulation mode is deterministic given the same timeline and clock progression, while generated responses remain model-dependent due to language-model inference and optional tool or memory routing. Emotional decay is applied during appraisal updates and PAD retrieval via timestamp-based decay terms. No resetting of internal state occurs between interaction steps within a run, allowing emotional and memory dynamics to accumulate over time.

B. *Emotion Model*

We adopt a hybrid emotion model that integrates appraisal theory with dimensional emotion representation (see Figure 1). Emotions are treated as constructed experiences when core affect is attributed to environmental events. Emotional appraisal is implemented as an explicit graph state that can be placed before reasoning (pre-appraisal), after reasoning (post-appraisal), or at both stages depending on agent configuration. At each appraisal stage, the model analyses the current message content, produces a PAD vector, and then uses an LLM call to estimate event intensity and importance. The resulting event is integrated into a bounded recent-event history with time-dependent decay, after which PAD is updated through a weighted shift toward the event vector. The updated PAD state is then re-injected into subsequent reasoning prompts (as either PAD values or textual affect descriptors), closing the perception, appraisal, and reasoning loop and enabling temporally continuous affect regulation across interaction steps.

Text-mediated Emotion Inference. Rather than mapping environmental sensor values directly to affective states, emotional inference is performed on the agent's generated responses. This design reflects our focus on emotion as an internal interpretative process rather than a direct function of environmental input. The same environmental stimulus can lead to different emotional responses depending on prior state, memory, and appraisal. By grounding emotional state updates in response text, we capture the interaction between perception, reasoning, and internal dynamics, and enable direct comparison between internal affect (PAD) and external expression (language), supporting the analysis of introspective affect reporting through internal–external alignment.

Dimensional Representation and Mapping. At any time t, the emotional state is a vector: $\mathbf{PAD}(t) = [P(t), A(t), D(t)]$.

The emotion model processes agent-generated responses and outputs probabilities over Ekman's six basic emotions using the Emotion English DistilRoBERTa-base model [30]. Let s_i denote the classifier score for emotion e_i on a message. PAD estimates are computed by a s-weighted sum of fixed PAD vectors for each emotion.

Temporal Dynamics. Emotional dynamics are modelled as an additive, decaying process:

$PAD(t) = \sum_{i=0}^{n} PAD_i \cdot e^{-\lambda(t-t_i)}$

with dimension-specific decay:

$P_{new} = P_{old} \cdot e^{-\lambda_P \Delta t}$

New events update the state as:

$PAD_{new} = PAD_{old} + (PAD_{event} - PAD_{old}) \cdot \omega(1 - \sigma)$

where ω is event intensity and σ is emotional stability. In practice, the transition to the new decayed target state is smoothed with a short sigmoid interpolation window to avoid abrupt jumps.

C. Experimental Setup

We introduce six agents with increasing architectural complexity, all operating on the same environmental input.

Agent Configurations. We start with a baseline agent, *Kevin*, which processes environmental events with minimal emotional processing overhead. *Trixie* applies appraisal before core reasoning. *Katya* applies appraisal after core reasoning. *Alex* combines pre-appraisal and post-appraisal stages. *Luka* extends dual-stage processing with stronger temporal continuity of affect. *Ian* adds emotional memory-centric mechanisms to the dual-stage architecture. Table I summarizes the agents.

All configured agents receive time and plant information tools, as well as live sensor access. Emotional agents additionally receive introspection tools, and memory-enabled agents receive retrieval/storage tools backed by vector memory.

TABLE I: Agent emotional processing capabilities

Agent	Pre-apr.	Post-apr.	Episodes	LT Memory
Kevin	×	×	×	×
Trixie	✓	×	×	×
Katya	×	✓	×	×
Alex	✓	✓	×	×
Luka	✓	✓	✓	×
Ian	✓	✓	✓	✓

Environmental Conditions. We evaluate agents under two controlled environmental scenarios: (1) a regular condition with stable and supportive environmental parameters, and (2) a stress condition introducing adverse factors such as delayed watering, suboptimal lighting, and temperature fluctuations. Each scenario consists of stimulated environmental sensor streams, derived from real-world data collected during prior deployments, creating a fixed sequence of environmental events (e.g., wake-up, humidity-drop, sleep), applied in identical order across agents. Agents maintain internal memories of prior appraisals and reactions, allowing past experiences to influence the interpretation of subsequent events and resulting in divergent behavioral trajectories over time. We call these conditions *a good day* and *a bad day* for the rest of the paper.

Input Conditions. We tested two different representations of emotional information; direct PAD values versus textual emotional descriptors, respectively. Across agents, this manipulation did not produce statistically significant differences, and we report only results using the latter.

IV. Experimental Results and Discussion

During runtime, the system stores per-step interaction artefacts generated within the execution pipeline, including environmental flags, generated messages, runtime, optional tool calls and responses, and emotional state snapshots (when enabled). To quantify computational and behavioural trade-offs, we aggregated $N = 3{,}443$ responses across all architectures.

We evaluated system behaviour across three dimensions: Internal–external emotional alignment and dynamics are evaluated by comparing internal PAD trajectories with external sentiment signals. External validation used VADER [31], [32], NRC-VAD [33], and CardiffNLP's Twitter-roBERTa-base [34] to analyse all responses. Environmental reactivity is assessed by measuring the magnitude and variance of emotional responses to environmental events across architectures. Behavioural consistency via topic modelling captures semantic structure using BERTopic [35] applied to the full response corpus, and is measured by the proportion of responses in which environment-derived topics are dominant. We also compare agent behaviours under normal (*good day*) versus stress (*bad day*) scenarios.

Internal-External Emotional Alignment and Dynamics. We analyse internal emotional dynamics and external textual expression separately before examining their alignment. Because each agent incrementally extends the previous architecture, these results can be interpreted as a controlled ablation over emotional-processing components. The separation

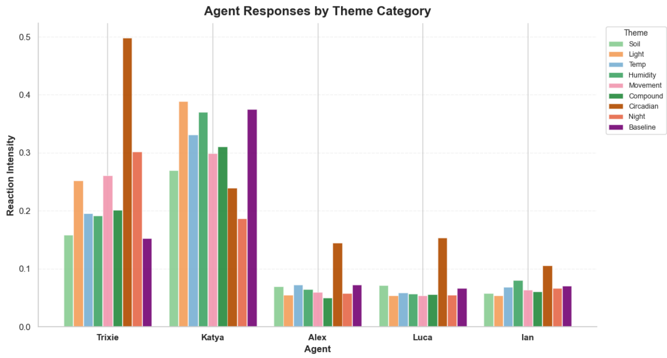


Fig. 2: Environmental stimulus impact and agent reactivity by architecture.

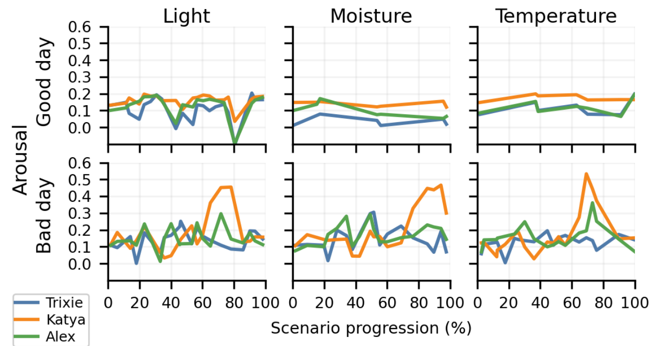


Fig. 3: Arousal progression across the top three event groups under good and bad day scenarios.

between internal dynamics and external expression enables a causal interpretation: internal emotional processes evolve continuously, while external outputs provide a discretised and partially filtered manifestation of these dynamics.

The key internal pattern remained consistent across complete architectures: pre-appraisal introduced negative PAD shifts, while post-appraisal recovered values toward baseline and stabilised final states reflecting the literature [14]. A controlled comparison of emotional input modalities (raw PAD values vs. textual descriptors) revealed no statistically significant population-level differences in sentiment distributions across agents (all $p > 0.05$, negligible effect sizes, Cramérś $V < 0.1$), despite observable local qualitative variation. Aggregate behavioural dynamics are largely invariant to representation modalities, and are instead driven by architectural factors such as appraisal structure. At the same time, local differences indicate that representation may still shape micro-level expression without altering global distributions.

Internal emotionality (PAD trajectories) and external textual sentiment showed architecture-dependent alignment. Lexical methods (VADER and NRC-VAD) showed high mutual agreement but strong positive skew, whereas CardiffNLP produced more balanced classifications, stability, and stronger alignment with PAD-derived labels. Overall agreement between internal and external emotionality remained moderate.

The moderate alignment across methods indicates that external sentiment measures capture only a partial projection of internal emotional state, highlighting a systematic gap between internal affect representation and linguistic realisation. This misalignment questions commonly used evaluation pipelines in multimodal interaction, which rely on observable signals such as language or expression: they may be systematically underestimating the complexity of internal affective dynamics. As a result, external behaviour alone may be insufficient as a proxy for emotional state in continuous, stateful agents.

Environmental Reactivity by Architecture. Environmental flags elicited distinct emotional response magnitudes by architecture. Figure 2 summarises both stimulus impact rankings and agent-level reactivity differences. Wake-up events produced the strongest average reactivity, followed by humidity-drop and sleep-transition events, indicating that temporal and moisture-related transitions dominate affective responses in this system. Agents are more sensitive to transitional environmental changes than to static conditions, emphasising the role of temporal events as primary drivers of affective response.

Agent-level reactivity exhibited a clear hierarchy. Katya (post-appraisal only) was most reactive, Trixie (pre-appraisal only) showed intermediate reactivity, and combined-processing agents (Alex, Ian, Luka) produced the most controlled and tightly clustered responses. This hierarchy is also visible in agent responses. Both excerpts below are responses to the same fungal-risk stimulus, allowing a direct comparison of how architecture shapes expression under identical environmental input. Under this stress event, Katya uses stronger emotional wording (e.g. "*the warm damp air feels heavy around me and I sense the risk of fungal growth creeping in*"; "*I need to stay vigilant…I hope the air dries out soon so I can thrive without worry* "). By contrast, Alex uses more practical, problem-solving language (e.g. "*my roots need good airflow to stay healthy*"; "*keeping the environment balanced will help me stay strong and resist any fungal issues*").

Behavioral Consistency. BERTopic analysis identified environmentally meaningful communication themes (e.g., light, stillness, moisture) across the full response corpus. Topic distributions differed significantly by architecture ($\chi^2 = 185.7$, $p < 0.001$), confirming emotional design choices produce distinct semantic behaviour signatures. Despite variation, all architectures preserved core task content. Emotional processing influenced regulation properties without suppressing relevant environmental awareness.

Cross-condition Comparison. Comparing the *good day* and *bad day* timelines provides insight into how the agents' behaviour reflects the impact of their emotional models under normal and stress conditions over time. Under stress, agents with single-stage architectures (specifically Katya/post-appraisal) exhibit significantly stronger and more volatile emotional responses. Pre-appraisal agent (Trixie) shows intermediate responsiveness, reflecting earlier processing without large amplification. Conversely, dual-stage architectures (Alex) clearly constrain response magnitude and variance.

TABLE II: Full event-indexed cross-scenario timeline showing all 6 environmental transitions and the comparative qualitative response patterns for Trixie, Katya, and Alex.

Event	Good-day stimulus	Bad-day stimulus	Trixie (pre)	Katya (post)	Alex (dual)
E1	Wake-up transition	Wake-up transition under stress context	Intermediate reactivity	High reactivity	Regulated reactivity with recovery
E2	Light available	Suboptimal/absent light	Intermediate salience shift	High salience shift	Lower-variance salience shift
E3	Timely watering	Delayed watering	Intermediate perturbation	Strong perturbation	Damped perturbation
E4	Humidity-drop transition	Humidity-drop under adverse context	Intermediate reactivity	High reactivity	Constrained reactivity
E5	Stable temperature	Temperature fluctuation	Intermediate PAD drift	Strong PAD drift	Reduced drift with stabilization
E6	Sleep transition	Sleep transition after stress accumulation	Intermediate closure response	Strong closure response	Controlled return toward baseline

They maintain stability across transient triggers (e.g., wake-up events) and sustained stressors (e.g., suboptimal lighting or temperature fluctuation). Furthermore, dual-stage agents demonstrate a faster, more regulated recovery towards baseline levels following cumulative stress during the sleep transition.

Emotional events form structured sequences where PAD states decay and accumulate dynamically, creating distinct context dependent trajectories based purely on architecture. Consequently, an agent's behaviour consists of continuous, state-driven interactions with environmental perturbations rather than discrete prompt–response pairs. Figure 3 captures this in the temporal evolution of arousal across the top-3 most reacted-to event groups (Light, Moisture, and Temperature). The dual-agent architectures are calmer because the emotional update is split across time, with pre-appraisal shaping initial evaluation and post-appraisal integrating the final response. Complementing this temporal view, Table II provides an event-indexed comparison across the full timeline, showing how these architectural differences manifest consistently across all environmental transitions. Importantly, these differences are not tied to specific environmental events, but emerge consistently across conditions, indicating that emotional architecture shapes the form of agent behaviour rather than its content. Emotional processing acts as a structural constraint on interaction dynamics, independent of the specific stimuli encountered.

Emotional Architecture as Behavioral Regulation. Across all analyses, the same pattern emerges: increasing emotional-processing complexity introduces measurable computational overhead, but in return yields more stable and coherent agent behaviour. Latency increased predictably with architectural complexity, but this cost was offset by improved temporal stability and more consistent responses. Post-appraisal proved to be the primary stabilising mechanism within the agent. Combined appraisal stages produced substantially more controlled environmental reactivity than single-stage designs, also at the event level (Figure 3), where dual-appraisal agents exhibit damped responses and faster recovery across identical event sequences under varying conditions.

The findings demonstrate a shift from reactions triggered by external events toward responses that are internally mediated as emotional-processing sophistication increases. This reframes emotion not as a discrete classification problem but as a continuous regulation process embedded within environment and agent interaction, where improved behaviour corresponds to more stable internal affect trajectories and their consistent externalisation.

V. Conclusion

It's time to grow, to drink in the sun, and to feel the world pulse gently with life around me. — Ian run excerpt

Our work addressed one of the core challenges in embodied artificial intelligence: enabling agents to process emotions while remaining grounded in environmental interaction. We developed plant monitoring agents that integrate PAD-based emotional modelling with continuous environmental sensing, demonstrating that emotional processing can be incorporated without compromising core monitoring functionality.

A central empirical finding is that dual-appraisal architectures provide substantially more stable emotional behaviour than single-appraisal systems. We observed that internal emotional states are not directly mirrored in language output, reflecting a more realistic but methodologically challenging model of emotional expression, revealing a systematic gap between affective state and its external expression. This also challenges the core assumption that observable signals are reliable proxies for internal states. Our proposed solution is to introduce evaluation frameworks that explicitly account for this decoupling and allow agents to explicitly represent and report their internal affective state. Additionally, we claim that structured emotion models can support more predictable and temporally coherent interactions, where improved behaviour corresponds to the stabilisation of internal affective state trajectories over time rather than purely surface-level response regularity.

Future work should extend this paradigm toward richer multimodal sensing and expression, tighter integration of memory and appraisal mechanisms, and interactive settings where environment–agent, human–agent, and agent–agent dynamics intersect, while incorporating human assessment of internal and external states into the evaluation framework, thereby enabling systematic study of how emotional regulation and internal states shape behaviour in real-world, continuous interaction scenarios.

REFERENCES

[1] Y. Li, Q. Sun, M. Schlicher, Y. W. Lim, and B. W. Schuller, "Artificial emotion: a survey of theories and debates on realising emotion in artificial intelligence," *arXiv preprint arXiv:2508.10286*, 2025.

[2] P. Fung, Y. Bachrach, A. Celikyilmaz, K. Chaudhuri, D. Chen, W. Chung, E. Dupoux, H. Gong, H. Jégou, A. Lazaric, *et al.*, "Embodied ai agents: Modeling the world," *arXiv preprint arXiv:2506.22355*, 2025.

[3] T. M. Moerland, J. Broekens, and C. M. Jonker, "Emotion in reinforcement learning agents and robots: a survey," *Machine Learning*, vol. 107, no. 2, pp. 443–480, 2018.

[4] A. Mehrabian, *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies*. Oelgeschlager, Gunn & Hain, 1980.

[5] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current psychology*, vol. 14, no. 4, pp. 261–292, 1996.

[6] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[7] C. Breazeal, "Emotion and sociable humanoid robots," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 119–155, 2003.

[8] T. Gomi and J. Ulvr, "Artificial emotions as emergent phenomena," in *Proceedings of 1993 2nd IEEE International Workshop on Robot and Human Communication*, pp. 420–425.

[9] A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. Grosset/Putnam, 1994.

[10] R. J. Dolan, "Emotion, cognition, and behavior," *Science*, vol. 298, no. 5596, pp. 1191–1194, 2002.

[11] A. Bechara, H. Damasio, D. Tranel, and A. R. Damasio, "Deciding advantageously before knowing the advantageous strategy," *Science*, vol. 275, no. 5304, pp. 1293–1295, 1997.

[12] G. Loewenstein and J. S. Lerner, "The role of affect in decision making," in *Handbook of Affective Sciences* (R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, eds.), pp. 619–642, Oxford University Press, 2003.

[13] S. C. Marsella and J. Gratch, "Ema: A process model of appraisal dynamics," *Cognitive Systems Research*, vol. 10, no. 1, pp. 70–90, 2009.

[14] K. R. Scherer, A. Schorr, and T. Johnstone, *Appraisal Processes in Emotion: Theory, Methods, Research*. New York, NY: Oxford University Press, 2001.

[15] A. A. Salah and A. El Ali, "Affective user interfaces," in *Handbook of Human Computer Interaction*, pp. 1–32, Springer, 2025.

[16] J. L. Salmeron, "Fuzzy cognitive maps for artificial emotions forecasting," *Applied Soft Computing*, vol. 12, no. 12, pp. 3704–3710.

[17] B. Schuller, A. Mallol-Ragolta, A. P. n. Almansa, I. Tsangko, M. M. Amin, A. Semertzidou, L. Christ, and S. Amiriparian, "Affective computing has changed: The foundation model disruption," *NPJ Artificial Intelligence*, vol. 2, no. 1, p. 16.

[18] M. Croissant, M. Frister, G. Schofield, and C. McCall, "An appraisal-based chain-of-emotion architecture for affective language model game agents," *Plos one*, vol. 19, no. 5, p. e0301033, 2024.

[19] L. F. Barrett, *How emotions are made: The secret life of the brain*. Pan Macmillan, 2017.

[20] K. Friston, "The free-energy principle: A unified brain theory?," *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[21] A. Clark, "Whatever next? predictive brains, situated agents, and the future of cognitive science," *Behavioral and Brain Sciences*, vol. 36, no. 3, pp. 181–204, 2013.

[22] E. Fox, "Perspectives from affective science on understanding the nature of emotion," *Brain and neuroscience advances*, vol. 2, p. 2398212818812628, 2018.

[23] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett, "The brain basis of emotion: a meta-analytic review," *Behavioral and brain sciences*, vol. 35, no. 3, pp. 121–143, 2012.

[24] R. A. Brooks, "Intelligence without representation," *Artificial Intelligence*, vol. 47, no. 1-3, pp. 139–159, 1991.

[25] R. Pfeifer and J. Bongard, *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press, 2006.

[26] H. Yin, Y. Cao, B. Marelli, X. Zeng, A. J. Mason, and C. Cao, "Soil sensors and plant wearables for smart and precision agriculture," *Advanced Materials*, vol. 33, no. 20, p. 2007764, 2021.

[27] A. Banks, "Mqtt version 5.0," tech. rep., OASIS, 2019.

[28] LangGraph, "Langgraph documentation," 2026.

[29] Langchain, "Langchain," 2026.

[30] J. Hartmann, "Emotion English DistilRoBERTa-base." https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/, 2022.

[31] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, 2014.

[32] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Sebastopol, CA: O'Reilly Media, Inc., 2009.

[33] S. M. Mohammad, "Nrc vad lexicon v2: Norms for valence, arousal, and dominance for over 55k english terms," *arXiv preprint arXiv:2503.23547*, 2025.

[34] D. Loureiro, F. Barbieri, L. Neves, L. E. Anke, and J. Camacho-Collados, "Timelms: Diachronic language models from twitter," Apr. 2022. arXiv:2202.03829 [cs].

[35] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," Mar. 2022. arXiv:2203.05794 [cs].

APPENDIX

A. Emotional Embeddings

The following table defines the PAD (Pleasure, Arousal, Dominance) values associated with different emotional states recognized by the plant monitoring system. These values are used to modulate the agents' behavior and responses based on the inferred emotional context.

All agents have access to basic environmental tools (`sense_environment`, `get_time_now`, `get_plant_information`), while emotional agents additionally have access to introspective capabilities (`check_how_you_feel`).

Tools available to each agent can be found in the table below:

B. Plant Monitoring System Flag Mappings

We provides a comprehensive reference of all environmental flags used in the plant monitoring system and their corresponding natural language messages. These mappings are defined in the `flagger_helper.py` module and determine how environmental conditions are translated into textual stimuli for the agents.

The environmental flags are organized into several categories based on the type of environmental condition they represent:

- **Soil Moisture** Flags related to watering events and soil moisture levels (`wl`, `wh`, `wat`)
- **Light Conditions** Flags monitoring light levels throughout the day (`ll`, `lmh`, `vhl`, `nlh`)
- **Temperature** Flags detecting temperature changes and extremes (`th`, `tl`, `tu`)
- **Humidity** Flags monitoring air moisture and humidity changes (`hl`, `hh`, `hs`, `hd`)

TABLE III: PAD (Pleasure, Arousal, Dominance) values associated with different emotional states recognized by the plant monitoring system.

Emotion	Pleasure	Arousal	Dominance
angry	-0.51	0.59	0.25
disgusted	-0.40	0.20	0.10
frightened	-0.64	0.60	-0.42
happy	0.46	0.20	0.15
sad	-0.4	-0.2	-0.5
surprised	0.30	0.67	0.25
bored	-0.65	-0.62	-0.33
curious	0.22	0.62	-0.01
dignified	0.55	0.22	0.61
elated	0.50	0.42	0.23
excited	0.76	0.67	0.35
inhibited	-0.54	-0.04	-0.41
loved	0.87	0.54	-0.18
puzzled	-0.41	0.48	-0.33
sleepy	0.20	-0.70	-0.44
unconcerned	-0.13	-0.41	0.08
violent	-0.50	0.62	0.38
neutral	0.0	0.0	0.0

TABLE IV: Agent tool configuration

Agent	Time	Sensors	Plant Info	Emotional State	Recent Memory	Search Memory	Emotional Search	Full Search	Store Memory
Kevin	✓	✓	✓	×	×	×	×	×	×
Trixie	✓	✓	✓	✓	×	×	×	×	×
Katya	✓	✓	✓	✓	×	×	×	×	×
Alex	✓	✓	✓	✓	×	×	×	×	×
Luka	✓	✓	✓	✓	×	×	×	×	×
Ian	✓	✓	✓	✓	✓	✓	✓	✓	✓

- **Interaction** Flags detecting human interaction and movement (`mv`, `nmv`)
- **Air Movement** Flags detecting drafts and air circulation (`dr`, `drs`)
- **Compound Conditions** Flags representing complex environmental states (`fr`, `lvpd`, `hvpd`, `or`)
- **General State** Flags representing periods of no significant activity (`huh`)
- **Circadian** Flags related to sleep-wake cycles (`sleep`, `wake`)
- **Nighttime Events** Flags for unusual nighttime activities (`nwat`)

C. Hyperparameter Configuration

The emotional and memory dynamics are governed by a fixed set of hyperparameters held constant across all agents to isolate the effect of architectural variation, ensuring comparability. These include emotional stability $\sigma = 0.5$, base emotional decay rate $\lambda = 0.1$, and an episode importance threshold (0.6), which determines when sustained emotional episodes are formed. Memory summarisation is controlled through a context-length threshold (default: 16 messages), after which recent context is truncated to a smaller retained window (default: 8 messages). The reasoning module temperature is configuration-driven (set to 1 in the current repository configuration). These parameters define the temporal sensitivity, persistence, and regulation strength of the emotional and memory systems. They were selected to ensure stable yet responsive emotional dynamics, avoiding both excessive volatility and overly dampened affective responses.

D. Prompts

This section provides the prompts used to generate the agent responses in the plant monitoring system. The prompts are designed to elicit specific emotional responses based on the environmental conditions and the agents' emotional states. Figure 4 is the prompt used for LLM-based emotional analysis in the plant monitoring system, and Figure. 5 is the reasoning prompt.

E. Performance and Computational Trade-offs

In addition to the three evaluations presented in the main paper, performance and computational trade-offs are assessed through runtime and response length in this section.

Emotional agents exchange computational speed for behavioural consistency, with an observed 87% overhead compared to baseline systems. Across $N = 3{,}443$ responses, runtime increased predictably with emotional-processing complexity. The observed differences correspond to moderate effect sizes, indicating that the trade-off between latency and stability reflects a meaningful behavioural shift rather than marginal statistical variation. Increased architectural complexity does not merely add processing overhead, but restructures

TABLE V: All environmental flags used in the plant monitoring system and their corresponding natural language messages.

Flag Code	Natural Language Message	Category
`wl`	soil's getting a bit dry. Should probably get some water.	Soil Moisture
`wh`	soil's pretty soggy. Might need to ease up on watering.	Soil Moisture
`wat`	someone just watered me. Nice and refreshing.	Soil Moisture
`ll`	it's pretty dark in here.	Light Conditions
`lmh`	i can feel the light. Just what I need for a productive day.	Light Conditions
`vhl`	ah, all this sunlight! Excellent for photosynthesis. Soaking it all up!	Light Conditions
`nlh`	why is it so bright? It should be dark now.	Light Conditions
`th`	getting a bit too warm for comfort here.	Temperature
`tl`	it's chilly. Everything's slowing down.	Temperature
`tu`	temperature's shifting a bit. Nothing major though.	Temperature
`hl`	air's feeling pretty dry.	Humidity
`hh`	it feels so humid I can almost swim through the air.	Humidity
`hs`	sudden humidity spike.	Humidity
`hd`	humidity just dropped.	Humidity
`mv`	i am being moved or at least i sense someone directly interacting with me.	Interaction
`nmv`	someone's moving around at night. Not the usual schedule.	Interaction
`dr`	is that a draft? I can feel the air moving.	Air Movement
`drs`	the draft stopped. The air is still again.	Air Movement
`fr`	warm and damp... not ideal. This could lead to fungus issues.	Compound Conditions
`lvpd`	air's pretty saturated. Transpiration's going to be slow today.	Compound Conditions
`hvpd`	dry air's making me lose water faster.	Compound Conditions
`or`	there seems to be a conflict between soil moisture and air humidity levels.	Compound Conditions
`huh`	nothing much happening. Just another quiet moment.	General State
`sleep`	im feeling tired. Time to slow down for the night.	Circadian
`wake`	i feel the morning light. Time to wake up from those strange plant dreams. i think i remember that i drempt about ...	Circadian
`nwat`	water? At this hour? Unusual timing.	Nighttime Events

```
You are tasked with analyzing the emotional content of the following text:
"{input_text}"

For the emotionality you will be given PAD values that come from the text of the
current event.
The current PAD values are:

- Pleasure: {pad_values["pleasure"]}
- Arousal: {pad_values["arousal"]}
- Dominance: {pad_values["dominance"]}

{If recent_occurrence is provided:}
    Recent context to consider: {recent_occurrence}

Analyze the text and return a JSON response that includes:
1. An intensity value between 0 and 1 - based on how intense the emotions are
2. An importance value between 0 and 1 - based on how important the event is to you
3. A brief emotional description (1-3 words)

Return strictly in this JSON format:
{
    "intensity": <float>,
    "importance": <float>,
    "emotional_description": "Max 3 words",
}
```

Fig. 4: LLM Emotional Analysis Prompt

the temporal dynamics of agent responses toward greater predictability.

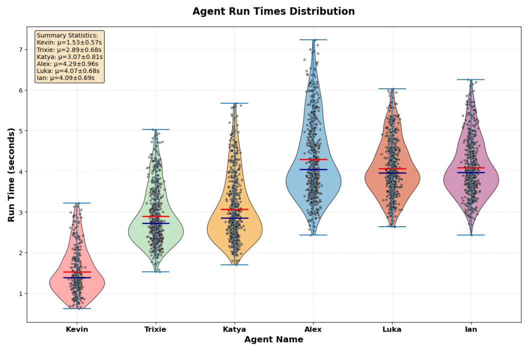
The baseline agent Kevin established the fastest response profile ($M = 1.92$s, $SD = 3.27$, $95\%\, CI\, [1.55, 2.28]$, $n = 313$). Basic emotional processing (Trixie) increased latency to $M = 3.58$s ($n = 626$), representing an 87% overhead relative to baseline. Advanced architectures (Alex, Ian, and Luka combined) reached $M = 4.60$s ($95\%\, CI\, [4.46, 4.74]$, combined $n = 1{,}878$), adding a further 28.4% beyond basic emotional processing. In practical terms, this creates a clear deployment trade-off. For 1,000 daily interactions, Kevin requires roughly 32.0 minutes of processing time, while emotional architectures require approximately 76.7 minutes.

```
You are a {plant_type} experiencing life in your environment.

You have access to various capabilities to sense and understand your surroundings.
"""

if self.profile.emodel:
    if self.profile.pad_text:
        system_message += """
        Current state: {emotionality}
        """
    else:
         system_message += """
         Internal state (PAD): {emotionality}
       """
    if self.profile.emotional_episodes:
        system_message += """
        Current context: {current_episodes}
        """

if self.profile.chroma:
    system_message += """
You have access to your stored memories and can recall past experiences
using the following abilities:
- Recall recent interactions to understand what has happened lately.
- Remember specific details or events from your memory.
- Search your memories by emotional content or similarity.
- Memorize new experiences or information for future recall.

Use these memory abilities when you need to reference the past, search for
emotional context, or store new information.
"""

system_message += """

Respond naturally based on your current situation and available information.
Never respond with numbers or code.
Respond based on all your knowledge and experiences in natural language.
"""
```

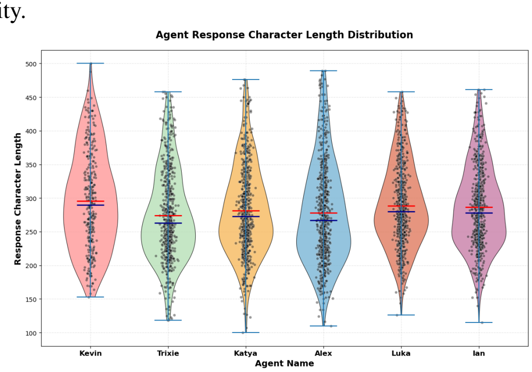
Fig. 5: Reasoning Prompt

However, this added latency is accompanied by improved temporal regularity, with advanced emotional agents exhibiting lower variability and more predictable behaviour than the baseline architecture.

Response-length analysis follows the same design logic. Kevin produced the longest responses on average ($M =$



(a) Agent runtime distributions by architecture complexity.



(b) Response character-length distributions across agent types.

Fig. 6: Performance and response characteristics across agents.

296.5 characters, $SD = 76.0$, $95\% CI\,[288.1, 305.0]$, $n = 313$), while emotional agents produced slightly shorter but more regulated outputs ($M = 278$–294 characters, combined $n = 3{,}130$). Overall, the baseline maximised raw throughput, whereas emotional architectures prioritised consistency and control. Figure 6 visualises these runtime tiers directly and confirms that latency scales with architectural complexity. Figure 6b shows the corresponding response-length distributions. Emotional architectures implicitly regulate output verbosity, reducing variance without enforcing rigid constraints on expression length.