International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 11 | Nov - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Adaptive Non-Player Characters Using Hybrid AI for Immersive Gaming

Chinmay Sharma
Student, AIT-CSE
Chandigarh University
Mohali-140413, Punjab, India
chinmaysharma2266@gmail.com

Yash Pathak
Student, AIT-CSE
Chandigarh University
Mohali-140413, Punjab, India
yashpathak0322@gmail.com

Utkarsh Yadav
Student, AIT-CSE
Chandigarh University
Mohali-140413, Punjab, India
utkarshy075@gmail.com

Yagnesh Reddy
Student, AIT-CSE
Chandigarh University
Mohali-140413, Punjab, India
vattiguntayagneshreddy@gmail.com

Pramod Vishwakarma
Associate Professor, AIT-CSE
Chandigarh University
Mohali-140413, Punjab, India
pramod.e9758@cumail.in

***Abstract*—Non-player characters (NPCs) remain disruptive to immersion through their repetitive speech patterns, failure to keep track of player behaviors, and uniformity in response across a variety of situations. Generative language models (LLMs) facilitate dialogue that is unplanned; yet, mere text prediction cannot ground an agent to its engine state, nor give it memory, disposition, and motivation. The paper suggests a hybrid architecture that consists of a biased but not superseded by an online player model (Bayesian Beta traits with forgetting, and role-pooled LinUCB bandit), and typed episodic–semantic memory, OCC/PAD emotions, and LLM dialogue layer constrained by schemas and validated via a permission system. The engine-less C# implementation in Unity, accompanied by 105 unit tests, is published. A simulation study carried out with this implemented system shows that forgetting lowers trait estimation error upon a change of player's style from 0.33 to 0.07; pooling the bandit across NPCs during one session regains 88% of gain that could be gained from an oracle (instead of 71% gain per each NPC), and keeping retrieval recency slots allows the most recent as well as the oldest memories to be preserved, outperforming all configurations of the recency half-life. The cognition process for eight NPCs takes 6.7 μs per frame. Within-subject user study (N=52, input modality-matched) is planned but not yet carried out.**

***Index Terms*—non-player characters, hybrid AI, game AI, player modelling, episodic memory, computational emotion, large language models, immersion, Unity**

I. Introduction

NPCs help make the virtual world feel occupied, but behavioral complexity has not matched the improvements seen in rendering and physics. Players still interact with characters that give a one-off greeting, have no memory of being attacked or rescued, and react without taking context into account—problems that have been extensively analyzed as immersion-breaking [1], [2]. Traditional NPC frameworks such as finite-state machines, behavior trees (BTs) [3], [4], GOAP [5], and utility-based choice [6] provide simple and authorable NPCs but do not respond to the unique player; if there is adaptation, it will be limited to difficulty adjustment [7].

Player modelling [8] tries to construct an estimate of the player based on telemetry, while large language models (LLMs) [9] enable practically limitless dialogue [10], [11]. But neither approach alone leads to believable, controllable, and grounded NPCs. A player model without a behavioral layer ends up producing no noticeable adaptation, while an unconstrained LLM may produce false scene states, switch back to assistant mode, fail to execute verified engine actions, and introduce latencies on the order of hundreds of milliseconds per query (compared to a 16.7 ms frame time at 60 FPS) [12], [13].

The proposed solution is based on the hybrid split approach: the deterministic utility selector works over manually designed behavior fragments to ensure that all physical actions stay within the designer-defined range, whereas the four auxiliary components take care of the profiling, remembering, affect, and conversation aspects. Specifically, the key contributions are:

1) **Veto-proof utility aggregation (Sections III–VI):** all learned biases in traits and bandits, affect in PAD terms, and dialogue intents go into the equation (1) as the capped multiplicative factors (maximal value of +50%) such that any non-zero factor vetoing the action results in the vetoing of the whole action.
2) **Engine agnostic C# 9 runtime (Section VII)** implemented as a Unity package, checked with 105 NUnit unit tests and working at 6.7 μs/frame with K=8 characters.
3) **Simulation study of the deployed code (Section VIII)** that gave rise to three design principles: to forget towards the prior when estimating the traits, to aggregate the strategy bandit across NPCs, and to save the memories in slots.
4) **User study protocol (Section IX)** that is ready for pre-registration and standardizes the input modalities between conditions and is powered to account for the

multiple comparisons.

Status. The architecture and runtime have been implemented and tested, but not the Unity scenario content and the user study; no participant data are available yet. All numeric results reported herein were generated with the released code.

II. Related Work

Scripted control and believability. FSMs, BTs [3], [4], GOAP [5], and utility-based AI [6] take just microseconds to compute and are easy to audit, but they treat the player as a collection of tuned variables. The Oz project [14] made believability their priority, and Fac¸ade [15] showed how authorial control and generative freedom had to be designed together.

Computational emotion. OCC [16] translates evaluation of events, agents, and objects into emotion labels. EMA [17] extends OCC with coping strategies, and ALMA [18] links OCC with PAD moods [19] inside a framework of OCEAN personality traits [20]. GAMYGDALA [21] packages OCC appraisal into a game engine module and leaves behavior and dialogue to the host game. Here we adopt ALMA and pass its results into utility functions, animation, and prompts.

Player modelling. Player taxonomies such as Bartle [22], or SDT [23] guide game difficulty or branching logic, while linear contextual bandits (LinUCB) [24] deal with exploration in fixed $O(p^2)$ time per update. Since end-to-end DRL policies preclude inspection and impose limits on safety constraints, online adaptation works with $d = 8$ Beta trait distributions and 13-parameter per role bandits.

LLM agents in games. Park et al. [10] rank free-form memory entries based on recency, importance and cosine similarity with embeddings, and use LLM reflection at regular intervals. Lyfe Agents [25] conserve tokens using option picking and summarization/forgotten storage; Voyager [26] compiles reusable skill code. Closely related to our approach is CICERO [27], where the strategic solver first commits to intents and then conditions dialogue generation. Survey work by Gallotta et al. [12] singles out grounding, persona shift, latency, and cost as challenges to game LLMs. In the blind ratings of Disco Elysium, Akoury et al. [13] report preference of players for lines written by studios over GPT-4 infills in 2:1 proportion whenever completion contradicted world facts or character's voice. Retrieval-augmented generation (RAG) lookup [28] lowers rate of fabrication but does not eliminate it [29].

Positioning. Generative-agent loops process every tick through LLM [10]; voice middleware [11] delegates non-verbal control to the host game scripts; GAMYGDALA [21] and separate player models target each faculty individually. We combine utility/BT control flow, online Beta/LinUCB player modeling, 256-slot typed memory, ALMA affect, and schema-checked dialogue into one veto-resilient loop; open source the C# implementation, and validate each subsystem via Monte Carlo simulation before human experiments.

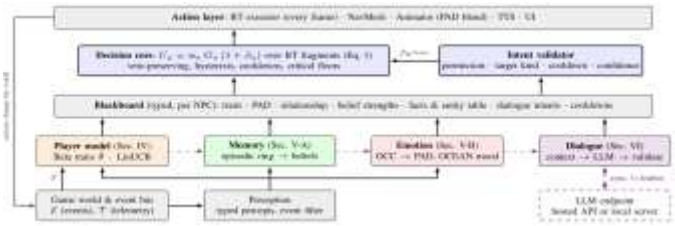

Fig. 1. Architecture of one hybrid NPC. Perception is the only reader of the world; four cognition modules write a typed blackboard; the decision core is the only writer of behaviour. Dialogue reaches behaviour solely as validated intents that scale utilities (Eq. 1). The player model is shared by all NPCs; bandits are shared per role. Dashed grey arrows are lateral reads (e.g., emotion reads retrieved memories).

III. Hybrid Architecture

A. Setting

In this work, we consider the setting of a single player game in a three-dimensional world with up to $K \leq 8$ NPCs present over 20–40 minutes long session. Virtual world produces an event stream E (e.g., PLAYERATTACKED, ITEMGIFTED, DIALOGUETURN, ...) and telemetry data T. For each NPC k, the problem is to choose a controller which selects actions from a designer-authored fragment set A_k and produces an utterance during dialogue turns. There are three restrictions on the system: authorial control (no actions outside A_k), real-time processing (non-dialogue cognition for all K NPCs in 0.5 ms from 16.7 ms frame) and safety (in-character, rating-compliant output with an authored fallback).

B. Layers

Figure 1 depicts the data flow. Each module supports Observe, Tick, Save/Load and can be disabled; if the module is disabled, it emits neutral trait/PAD values (0.5) and neutral belief values (0), thus making it clear that ablation always comes down to a configuration modification.

Cognition proceeds in round-robin fashion, updating $\lceil K/4 \rceil$ NPCs per frame and thus guaranteeing that each NPC is updated at least once per four frames (at least 67 ms at 60 Hz), whereas BTs and reflex guards are updated every frame.

C. Utility with Veto-Preserving Bonuses

Each fragment a consists of a set of considerations C_a, each of which is an input $x_c \in [0, 1]$ fed into a designer-shape response function $f_{a,c}$ (can be linear, polynomial or logistic [6]). Utility of fragment a is defined as follows:

$$U_a = \max\left(w_a G_a (1 + \bar{\beta}_a), u_a^{\text{floor}}\right) \tag{1}$$

$$G_a = \prod_{c \in \mathsf{C}_a} f_{a,c}(x_c)^{1/|\mathsf{C}_a|}. \tag{2}$$

where G_a is the geometric mean of the consideration scores, $\bar{\beta}_a = \min(\beta_a, 0.5)$ caps $\beta_a \geq 0$, the summed bonus from validated dialogue intents and the bandit's current strategy (Sec. IV), and u_a^{floor} is nonzero only for quest-critical fragments. G_a keeps large consideration sets from collapsing the score,

and because the bonus multiplies it, any consideration at zero vetoes the fragment. A dialogue intent can tip a close decision, but it cannot make a terrified NPC follow the player into danger. (With an additive bonus, $G_a = 0$ would still score β and could win; a unit test covers this case.)

The running fragment gets a 10% hysteresis bonus, and the others respect cooldowns. The scripted baseline of Sec. IX drives the same fragments with an FSM, so action, animation and audio are shared.

IV. Learning Player Preferences

Within minutes and without the aid of any labels off-line, it is imperative for the player model to develop an interpretive form which can be read by game designers, used by utilities, and paraphrased by prompts. This process is divided into two parts: the first being trait assessment (what the player does), and the second being strategy choice (what the NPC tries to do).

A. Traits

There are eight propensity dimensions being recorded here with $d = 8$ on $[0, 1]$ namely: aggression, mercy, exploration, sociability, pace, achievement, generosity and humor. Each trait j has evidence events (e.g., ENEMYSPARED gives mercy observation $o = 1$) and a Beta posterior that is updated after every observation, with forgetting toward the prior $(\alpha_0, \beta_0) = (2, 2)$:

$$\alpha \leftarrow \lambda\alpha + (1 - \lambda)\alpha_0 + o, \tag{3}$$

$$\beta \leftarrow \lambda\beta + (1 - \lambda)\beta_0 + (1 - o), \tag{4}$$

$$\hat{\vartheta} = \frac{\alpha}{\alpha + \beta}. \tag{5}$$

For $\lambda = 0.98$, the sample size saturates to $\alpha_0 + \beta_0 + 1/(1 - \lambda) = 54$, and the estimator keeps tracking the drift (Section VIII). Slow decay in time, with a half-life of 10 minutes, pulls the unseen traits closer to the prior than locks down the outdated information.

The utilities and the bandit use an "attenuated" trait

$$0.5 + c_j(\hat{\vartheta}_j - 0.5), \tag{6}$$

where

$$c_j = 1 - \min(1, \sigma_j/\sigma_0) \tag{7}$$

is the posterior standard deviation divided by the prior, guaranteeing neutrality in the cold start period.

Traits for the prompts only are soft-mapped into the designer archetypes (Warrior, Diplomat, Wanderer, Speedrunner) using a Gaussian kernel; a weak mapping results in a hedged assertion ("you are not sure what to make of the player").

B. Strategies as a Role-Pooled Contextual Bandit

Each NPC role has a few interaction strategies S (a merchant: banter first, straight to business, offer a rumour, request a favour). Given the context

$$x = [\hat{\vartheta}; \text{PAD}; r; 1] \in \mathbb{R}^{13} \tag{8}$$

(r is the relationship, Sec. V-A), disjoint LinUCB [24] picks

$$s^* = \arg\max_{s \in S} \left[w_s^T x + \alpha_{\text{ucb}} \sqrt{x^T A_s^{-1} x} \right], \tag{9}$$

$$w_s = A_s^{-1} b_s. \tag{10}$$

with $\alpha_{\text{ucb}} = 0.3$; we maintain A by Sherman–Morrison, so a decision plus update is $O(p^2)$ and allocation-free. Here, the reward $r \in [0, 1]$ works as the engagement parameter for the designer weighted over a period of 30 seconds, where the player was engaged in the conversation, opened up any extra topics, responded to the content, and did not skip anything.

The chosen strategy adds an effect of $\beta_a = 0.25$ to the fragments using that strategy and includes one line of direction in the prompt. One NPC provides around ten interactions, which are not enough to cover 13 parameters per arm, so all NPCs that share a strategy set share one bandit: within a session they pool evidence about the same player, and across sessions the bandit can persist as a population prior. Sec. VIII quantifies both effects.

V. Memory and Emotion

A. Typed Episodic–Semantic Memory

Memory is a typed record, not free text (cf. [10]), containing event type, participants, timestamp, importance $I \in [0, 1]$, valence, a 64-bit topic mask, and template summary. Language model (LLM) only reads summaries but never writes records. There is always a fixed-size per-NPC array with 256 slots of records without any dynamic allocation.

Importance I depends deterministically on event type, damage amount, quest state change, and rarity; records with $I < 0.15$ are not stored. Once the array becomes full, the record with the lowest joint measure of importance and recency gets evicted. A record m is ranked against a query q (situation tags, participants, and position) with the formula

$$\text{score}(m, q) = 2^{-\Delta t/h} + I + \text{rel}(m, q), \tag{11}$$

where Δt is the time since last access, h is recency half-life, and rel is a symbolic relevance term computed as 0.6 * Jaccard similarity of tags + 0.25 * number of common participants + 0.05 * number of secondary participants + 0.1 * spatial kernel.

Lookup scans the 256-record array with the help of bounded insertion sort. For dialogue prompt, the slot-reserved retrieval policy is applied: slots 1–3 unconditionally store three latest episodes to keep the context of the immediate turn, while slots 4–8 rank records according to Eq. 11 with $h = 40$ min. The reason behind this approach (illustrated in Fig. 3) is that there is no single decay rate in an additive score that would be able to maintain both recent turns and more distant salient episodes.

Once every 120 seconds (for the cost of 14.3 μs), the memory consolidation process forms semantic beliefs from all clusters of at least three episodes sharing the (type, subject) with strength $1 - e^{-(\sum I)}$ and importance-weighted valence. Semantic beliefs decay with 60 min half-life and apply utility curves to normalized $[0, 1]$ blackboard inputs.

Semantics beliefs exist above the raw episodes in the prompt list, therefore creating behavioral abstraction by Park et al. [10] for the price of zero LLM calls. For instance, merchant's REFUSETRADE fragment reads the belief "stolen from me" using the logistic curve. This way, memory is wired into the same toolset as any other kind of input. The relationship

$$r = \tanh\left(-\frac{1}{2}\sum_m I_m v_m\right) \tag{12}$$

$\in [-1, 1]$, which is defined over the records with the participation of the player, is exposed to both the utilities and the bandit.

B. Appraisal-Driven Emotion and Mood

Each NPC possesses an OCEAN vector $\psi \in [0, 1]$ [20]. Events have appraisal variables authored by the designer including desirability, praiseworthiness, probability, responsible agent and appeal, that map to twelve OCC categories [16]: joy/distress, hope/fear (prospects), admiration/reproach, gratitude/anger, pride/shame, and liking/disliking. Complex emotions consume their components.

In the case where the player causes damage to the NPC, anger consumes distress and reproach but does not compete with it. Impulse vectors are weighted by personality (increasing the intensity of negative emotions for neuroticism and decreasing anger for agreeableness) and projected into PAD space through ALMA's mapping [18], $W \in \mathbb{R}^{3\times 12}$ (e.g., anger $\mapsto$ $(-0.51, 0.59, 0.25)$).

Emotion e and mood m evolve as

$$e_t = \text{clip}\left(e_{t-1}e^{-\Delta t/\tau_e} + W\Delta e\right), \tag{13}$$

$$m_t = m_{t-1} + \frac{\Delta t}{\tau_m}(m_0 + \gamma e_t - m_{t-1}), \tag{14}$$

with $\tau_e = 5 + 55\psi_N$ s, $\tau_m = 120 + 780\psi_N$ s, $\gamma = 0.3$ and baseline mood $m_0(\psi)$ from Mehrabian's OCEAN-to-PAD regression, as in ALMA.

Eq. 14 relaxes exactly to m_0 once emotion has decayed. The more common form $\gamma e + (1-\gamma)m$ settles at $(1-\gamma)m_0$ instead, a bug our tests caught. The effective state $\text{clip}(e + m)$ drives utilities (FLEE rises with high arousal and low dominance), Animator blend parameters, TTS pitch and rate, and a single prompt line that names the dominant emotion category and its cause.

VI. Grounded LLM Dialogue

A. Contextual Grounding

Figure 2 charts the asynchronous turn loop. Prompt assembly packs symbolic blackboard state into a 1,200-token

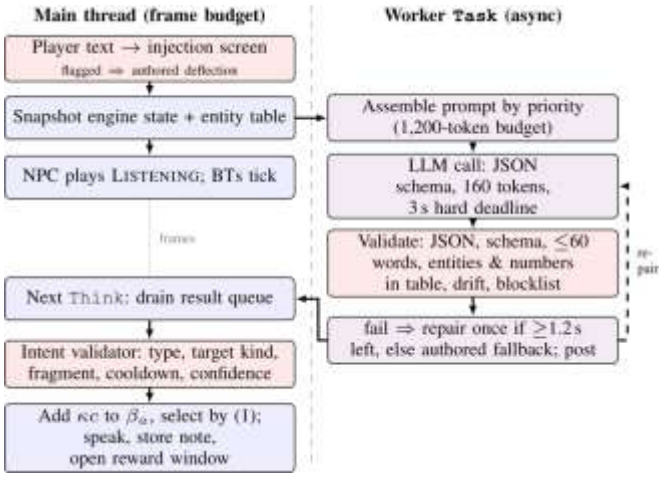

Fig. 2. Asynchronous dialogue turn (1,200-token cap, 3 s timeout). Shaded red blocks mark deterministic guard-rails; any check failure or timeout emits an authored fallback line. Validated intents scale Eq. 1 utilities by at most +50%.

window under a strict drop order: (1) persona card and JSON schema rules (pinned), (2) local scene facts, (3) semantic beliefs, (4) dominant OCC emotion and cause, (5) Gaussian archetype impression, (6) active LinUCB tactic, (7) the eight slot-reserved episodes from Sec. V-A, and (8) the last six dialogue turns. Overflow truncates from the bottom up. Because every field mirrors live engine state (symbolic RAG [28]) and appraisal is authoritative, the prompt prescribes the NPC's active emotion rather than letting the LLM invent one.

B. Output Formatting and Validation

Answers have to deserialize to a strict JSON schema: an utterance string ($\leq$ 60 words), up to three intents triples (type, target, confidence $\in [0, 1]$), and optional memory_note (when grammatical constraints can be decoded). Contextual prompts are not enough to stop output hallucinations, hence a deterministic C# validator (6.95 μs per reply) inspects each completion before playback.

Every capitalized substring in the middle of a sentence has to be matched with a scene entity or persona-lexicon term; every number larger than 10 has to be matched with a blackboard number (prices, quantities). Rating-blocklist hits and meta-game/assistant references like "as an AI" or "the player" are rejected as well.

In case of a validation failure with $\geq$ 1.2 s remaining until the deadline, the invalid JSON and corresponding error code are re-issued for one repair attempt; otherwise the authored line is replayed immediately. Player input matching any prompt-injection regexes is filtered out before sending the request.

C. Intents

Verbal acts influence the non-verbal performance of NPCs only through the intents array (0.23 μs check). An intent is valid if and only if the following four criteria are satisfied:

(i) its verb is on the allow-list of that NPC (trader allows OfferDiscount but not OpenVault), (ii) its target resolves to an existing scene entity of appropriate type, (iii) the corresponding BT fragment is defined and available, and (iv) confidence $c \geq 0.3$.

Each accepted intent contributes κc ($\kappa = 0.4$) to that fragment's multiplier in Eq. 1—tipping close utility tie-offs while retaining zero-consideration rejection.

D. Backends

There is one OpenAI-compatible HTTP client able to contact either cloud endpoints or locally hosted 8B Llama 3.1 weights [30]; our protocol benchmarking compares the two. Prompts contain only the coarse archetypes and trait buckets (never the raw telemetry).

VII. Implementation

Sec. III–VI modules along with the FSM base line and the NDJSON logger assemble in a stand-alone C# 9 executable (Unity 2021.3 LTS version) running without any engine dependencies.

The MonoBehaviour bridge is responsible for the $\lceil K/4 \rceil$ scheduling, updating the active BT nodes, and sending PAD float variables into the Animator blend trees.

Thanks to maintaining the engine-independent core design, the dotnet testing framework can run all 105 NUnit tests in headless mode, including the verification of Eqs. (1)–(6), all validator rejection codes, timeout and single-retry repair workflows (checked with a deterministic fake client), module ablation flags, and multi-step scenarios (like three theft cases building up into the belief "stole from me", triggering REFUSETRADE response).

Main thread workflows allocate zero managed memory heap space (by using value-type struct, fixed arrays and Span slices); HTTP I/O, NDJSON parsing and validation are done by the background Task workers, with the results forwarded via the lock-free concurrent queue.

VIII. Simulation Study

A. Measured CPU Cost

Table I reports the measured CPU cost (median of 201 batches) on AMD Ryzen AI 9 HX 370, .NET 8.0.31 JIT. Unity's Mono/IL2CPP backends will differ; these are reference values, not a console profile.

Table I uses $K = 8$ NPCs with 12 three-consideration fragments and full memories. A main-thread frame costs 6.7 μs, about 1% of the 0.5 ms budget, which leaves plenty of margin for slower backends. Retrieval (3.2 μs) is five orders of magnitude cheaper than an LLM call.

B. Simulation Setup

Before involving participants, we test each adaptive component against synthetic ground truth using the compiled runtime that ships in the package; seeds, data and Fig. 3 regenerate with one command.

TABLE I
Measured CPU Cost (Median of 201 Batches)

Operation	Unit	μs
Perception: one percept (memory store + appraisal)	1 percept	4.59
Player model: one trait observation	1 obs.	0.02
Emotion integration, one NPC	1 NPC	0.12
Utility selection, 12 fragments × 3 considerations	1 NPC	0.94
Full cognition tick (Think), one NPC	1 NPC	2.43
Behaviour-tree tick, all 8 NPCs	8 NPCs	0.27
Memory retrieval, top-5 of 256	1 query	3.18
LinUCB select + update, $p = 13$, 4 arms	1 decision	0.87
Reply parse + validation (off main thread)	1 reply	6.95
Intent validation and bonus	1 reply	0.23
Consolidation, 256 records	Every 120 s	14.3
Whole frame: 2 Think + 8 BT ticks + 1 percept	Frame, $K = 8$	6.7

C. Trait Estimation

Traits correspond to a different play style (Fig. 3a). The evidence event is represented by a Bernoulli observation of the corresponding propensity. With forgetting, the median estimate crosses 0.5 after 33 events since the switch, and the mean absolute error for the last 50 events is 0.07. Without forgetting, 133 events are needed (and only 48% of seeds cross), with the error of 0.33. The 10–90% confidence interval right before the switch covers 0.10, showing the price of a 54-event memory.

D. Bandit Pooling

Pooling increases the utility of the bandit in a single session (Fig. 3b). Players are simulated using four archetypes, and the engagement with each strategy is a fixed linear function of their true traits. However, the bandit sees only the shrunken estimates obtained from the player model with three observations per trait between interactions, where rewards have Gaussian noise ($\sigma = 0.15$).

During 40 interactions on average the bandit obtains the following rewards: 0.403 for random choice, 0.589 for a single bandit per NPC, 0.633 for a single bandit for the whole set of four NPCs, 0.646 when a population prior is used, and 0.665 for an oracle.

The bandit with a population prior closes 93%, the pooled one—88%, and the per-NPC one—71% of the gap to the oracle. The per-NPC bandit is worse than random at the beginning since it has to explore all the arms; pooling reduces the burden, and the population prior removes it almost completely.

The reward model is closed, so the ranking and the cold start penalty, but not the actual values, should be expected to work elsewhere.

E. Memory Retrieval

Retrieval needs reserved recency slots (Fig. 3c). The ranking using criterion 11 alone creates a dilemma: short half-lives increase the significance of the newest events while hiding the important ones from the past (0% at $h = 10$ min), and

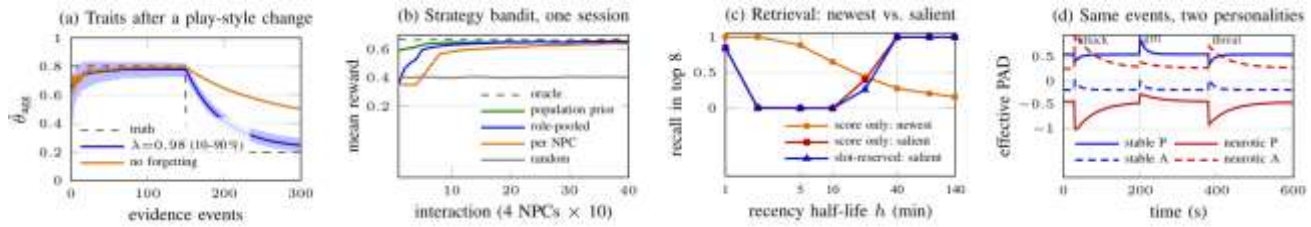

Fig. 3. Simulation study on the shipped runtime. (a) True aggression switches from 0.8 to 0.2 after 150 events; median and 10–90% band over 1000 seeds. (b) 2000 simulated sessions; role-pooled NPCs share one bandit per session; the population prior is also warmed on 50 earlier players. (c) 30-min sessions, 3 salient events in the first 10 min and a mundane event every 10 s; 1000 trials per point; dotted: chosen $h = 40$ min (slot-reserved newest-recall is 100% everywhere). (d) Pleasure (P) and arousal (A) of a stable, agreeable ($\psi_N = 0.15$) and a neurotic ($\psi_N = 0.85$) NPC.

long half-lives create the inverse situation (newest-three recall = 28% at $h = 40$ min).

Very short half-lives ($h = 1$ min) work quite well under such conditions only because recency acts like a one-minute window and importance ranks the rest, depending on the rate of events. Three slots allocated to the newest events and $h = 40$ min for the rest achieve 100% recall for both newest and important events.

F. Personality-Dependent Effects

Personality-dependent effects can be noticed (Fig. 3d). An identical attack decreases the pleasure by 0.15 in the case of a stable non-player character (NPC) and 0.63 in the case of the neurotic NPC, with the emotion having the half-decay time of 37 s compared to 10 s. The neurotic NPC is also characterized by stronger fear (higher arousal and lower pleasure) after the next threat.

G. Validator Sanity Check

We ran a sanity test for the C# validator against the 20-reply benchmark, consisting of 10 valid grounded responses and 10 invalid responses with invented town/NPC names, unobserved prices (¿10), assistant boilerplate, or meta-game jargon.

The result was: 10/10 valid replies passed; 10/10 faulty replies were properly rejected with the correct error code. In the early testing, a lowercase fourth-wall leak (“check your quest log”) bypassed entity capitalization checks, resulting in including meta-game phrases into the drift blocklist.

This 20-reply sample constitutes a deterministic regression test rather than the large language model hallucinations field estimate.

IX. User Study Protocol

Protocol specification below follows OSF pre-registration standards; recruitment awaits institutional ethics clearance (0 participants run to date).

A. Conditions and Content

In our crossover design, Scripted sequences the shared BT fragments via a conventional FSM and static dialogue tree, whereas Hybrid enables the full Sec. III–VI stack.

To avoid confounding cognition with input mode (menu clicks vs. typing), players use the exact same free-text box in both arms: under Scripted, a token-overlap matcher routes

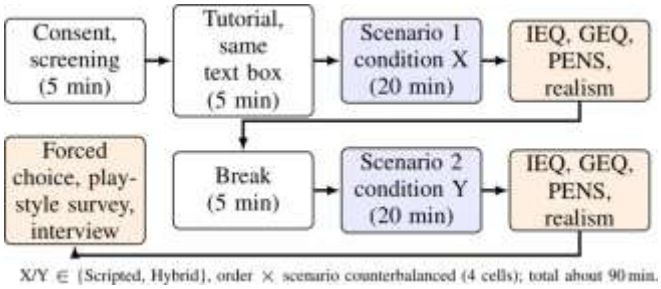

Fig. 4. Crossover session timeline (2 × 20 min). Both arms share identical free-text UI, 3D assets, TTS voices and BT fragments; only the cognitive controller changes.

typed input to the highest-overlap authored branch (handling paraphrases) or plays a fallback “didn’t catch that” line below threshold.

Each player completes two matched 20 min village hubs (populated by a trader, guard captain, innkeeper and child across three short quests) with condition sequence and village pairing counterbalanced (Fig. 4). Two external narrative designers, blind to H1–H5, authored the Scripted trees under a fixed 40 person-hour cap per village.

B. Sample Size and Analysis

Paired t-tests on hypotheses H1–H4 with Holm correction (minimum $\alpha = 0.0125$) need $N = 48$ subjects to detect an effect size of $d_z = 0.5$ with power 0.80 (non-central t distribution).

In all, $N = 52$ subjects were recruited to account for exclusions (failing the attention check and ¿10% technical loss), yielding power of 0.81 for hypothesis H5 (one-sided, margin $d_z = 0.5$).

Wilcoxon test will be used instead of t-test if Shapiro-Wilk test suggests a violation of normality, and effects will be described with d_z with 95% confidence intervals. A mixed-effects model will include order and scenario.

Because of possible within-subjects bias, which occurs because contrasts become exaggerated after participants go through the enriched condition, between-subjects contrast in the first period will also be conducted.

There will be five ablation conditions (–PlayerModel, –Memory, –Emotion, –LLM, full) with $N = 20$ subjects per

arm, which are labeled a priori exploratory and which are able to detect $d \geq 0.91$ with power 80%.

De-identified NDJSON logs, R/Python code, and Unity binaries will be uploaded on OSF.

C. Hypotheses

H1–H4 (primary, confirmatory) state that Hybrid will outperform Scripted in terms of IEQ immersion [1], GEQ flow [31], PENS relatedness [23], and 16 item scale of NPC realism (covering memory, emotions, conversation, and consistency dimensions; validated with 12 participants before the study for internal consistency, $\alpha \geq 0.70$).

H5 (non-inferiority) states that consistency score obtained under the Hybrid condition does not fall behind the one obtained under the Scripted condition by more than 0.5 on the scale, in order to address the consistency penalty of naive LLM conversations [13].

Additional variables include telemetry (optional NPC time, player-initiated turns, return rate, skipping rate), two uncanny valley tests [32], forced choice preference, latency, and fallback rate.

As the reward function of the bandit uses engagement telemetry as the criterion, telemetry outcomes are considered secondary. The validity of the player models is assessed through correlation of self-rated play style and the final estimate of traits $\hat{\vartheta}$.

X. Discussion and Conclusion

A. Why Hybrid

Episodic lookup, affective decay, player profiling, and surface realization function on different time scales (microseconds vs. seconds). For our architecture, scanning 256 entries of typed memories takes 3.18 μs and provides utility curves without verbalization; OCC/PAD fusion (0.12 μs) gives Animator/TTS float values rather than text; and Beta/LinUCB updates (0.89 μs) are easily inspected.

The big language model is called only for generating speech phrasing according to schema and entity limitations and allow-lists, so if there are generation failures, they are limited to one scripted line within the 3-second limit.

Since players dislike inconsistency between the dialogue and current state of the scene [13], calculation of utilities by Equation 1 and prompting is done by the same authoritative blackboard.

B. Scope of the Evidence

Monte Carlo experiments in Section VIII check Equations (1)–(6) on artificial data with 1,000–2,000 seeds and suggest three specific improvements: lambda 0.98 prior reverting Beta priors, LinUCB with pooled role parameters, and 3+5 slot reservations retrieval policy.

These experiments make use of stylized reward and event generators; nevertheless, the benefits of human immersion in terms of $N = 52$ crossover trial in Section IX is still a matter of empirical validation.

C. Limitations and Risks

The next section highlights the control parameters for interaction dynamics in our target system. Restrictions on intent bonuses ($\bar{\beta}_a \leq 0.5$) ensure initial dialogue (G1) but might make it mechanical in case of dominant base utilities; it is fine-tuned using κ (default value 0.4), alongside low-base optional fragments, such as OFFERDISCOUNT.

On the other hand, the interplay of memory, mood, and the bandit-driven transition dynamics might result in blocking required interactions (e.g., a displeased NPC refusing to participate). Utility floors ($u_a^{\text{floor}} > 0$) prevent this from happening on mission-critical fragments.

The surface validator rejects any unattested capitalized proper noun to enforce strong grounding at the expense of reduced lexical diversity; however, it fails to identify false predicates for valid entities (e.g., “Mira left the town”), thus providing us with residual error rate that will be logged during user trials.

Authoring cost per each NPC (personality, OCC, OCEAN personality, curves, allow-list) will be also logged. Finally, response latencies between 1 and 2 seconds are both deliberation and possible lag; expressive speech with static animation is risky due to uncanny valley effect [32]. Novelty effects are mitigated in part by the –LLM cell.

D. Conclusion

We have designed, tested (using 105 NUnit tests), and released a mixed NPC design that biases the deterministic utility/BT core solely using veto-proof multipliers through Beta/LinUCB profiling, a 256-slot episodic–semantic memory, ALMA affect, and LLM dialogue that is validated by schemas according to Equation 1.

In simulation experiments carried out on our delivered C# runtime, we have detected two issues with the formulation of this system, as well as prior-reverting Beta tracking (mean absolute error 0.07 compared to 0.33), role-pooled bandits (88% compared to 71% oracle gain), and 3+5 slot-reserved retrieval (100% recall) at 6.7 microseconds per frame for $K = 8$ NPCs.

The planned next steps are to profile the Unity build on study hardware, pilot the 16-item realism scale ($N = 12$), benchmark hosted versus local Llama 3.1 8B backends, and run the pre-registered $N = 52$ crossover study.

References

[1] C. Jennett et al., “Immersion experience measurement in digital games,” *Int. J. Hum.-Comput. Stud.*, vol. 66, no. 9, pp. 641–661, 2008.
[2] E. Brown and P. Cairns, “Grounded study of game immersion,” in *Proc. ACM CHI Ext. Abstr.*, 2004, pp. 1297–1300.
[3] D. Isla, “Managing Halo 2 AI complexity,” in *Proc. GDC*, San Francisco, CA, 2005.
[4] M. Colledanchise and P. Ögren, *Behavior Trees in Robotics & AI*. Boca Raton, FL: CRC, 2018.
[5] J. Orkin, “Three states and plan: F.E.A.R. AI,” in *Proc. GDC*, San Jose, CA, 2006.
[6] D. Mark, *Behavioral Math for Game AI*. Boston, MA: Cengage, 2009.
[7] R. Hunicke, “Dynamic difficulty adjustment in games,” in *Proc. ACM ACE*, 2005, pp. 429–433.

[8] G. N. Yannakakis et al., "Player modeling," in *Artif. Comput. Intell. Games*, vol. 6. Dagstuhl: LZI, 2013, pp. 45–59.
[9] T. B. Brown et al., "Language models as few-shot learners," in *Proc. NeurIPS*, vol. 33, 2020, pp. 1877–1901.
[10] J. S. Park et al., "Generative agents: Simulacra of human behavior," in *Proc. ACM UIST*, 2023, pp. 1–22.
[11] NVIDIA Corp., "ACE avatar cloud engine," 2023. [Online]. Available: https://developer.nvidia.com/ace-for-games
[12] R. Gallotta et al., "LLMs and games: Survey and roadmap," *IEEE Trans. Games*, pp. 1–18, 2024.
[13] N. Akoury et al., "Player perceptions of LLM game dialogue," in *Findings ACL: EMNLP*, 2023, pp. 2295–2311.
[14] J. Bates, "Emotion in believable agents," *Commun. ACM*, vol. 37, no. 7, pp. 122–125, 1994.
[15] M. Mateas and A. Stern, "Interactive drama in Fac¸ade," in *Proc. GDC*, 2003.
[16] A. Ortony et al., *Cognitive Structure of Emotions*. Cambridge, U.K.: CUP, 1988.
[17] S. Marsella and J. Gratch, "EMA appraisal dynamics model," *Cogn. Syst. Res.*, vol. 10, no. 1, pp. 70–90, 2009.
[18] P. Gebhard, "ALMA layered affect model," in *Proc. AAMAS*, 2005, pp. 29–36.
[19] A. Mehrabian, "PAD framework for temperament differences," *Curr. Psychol.*, vol. 14, no. 4, pp. 261–292, 1996.
[20] P. Costa and R. McCrae, *NEO-PI-R and NEO-FFI Manual*. Odessa, FL: PAR, 1992.
[21] A. Popescu et al., "GAMYGDALA emotion engine," *IEEE Trans. Affect. Comput.*, vol. 5, no. 1, pp. 32–44, 2014.
[22] R. Bartle, "Hearts, clubs, diamonds, spades: MUD players," *J. MUD Res.*, vol. 1, no. 1, 1996.
[23] R. Ryan et al., "Motivational pull of games via SDT," *Motiv. Emot.*, vol. 30, no. 4, pp. 344–360, 2006.
[24] L. Li et al., "Contextual-bandit news article recommendation," in *Proc. WWW*, 2010, pp. 661–670.
[25] Z. Kaiya et al., "Lyfe agents for real-time social interactions," arXiv:2310.02172, 2023.
[26] G. Wang et al., "Voyager: Embodied agent with LLMs," arXiv:2305.16291, 2023.
[27] Meta FAIR Diplomacy Team et al., "Human-level Diplomacy via LLMs and planning," *Science*, vol. 378, no. 6624, pp. 1067–1074, 2022.
[28] P. Lewis et al., "RAG for knowledge-intensive NLP tasks," in *Proc. NeurIPS*, vol. 33, 2020, pp. 9459–9474.
[29] Z. Ji et al., "Hallucination survey in natural language generation," *ACM Comput. Surv.*, vol. 55, no. 12, pp. 1–38, 2023.
[30] A. Grattafiori et al., "Llama 3 herd of models," arXiv:2407.21783, 2024.
[31] W. IJsselsteijn et al., "Game experience questionnaire," Tech. Univ. Eindhoven, Rep., 2013.
[32] M. Mori et al., "The uncanny valley," *IEEE Robot. Autom. Mag.*, vol. 19, no. 2, pp. 98–100, 2012.