FROM EXPERIENCE TO DELIBERATION: AN ADAPTIVE COGNITIVE HARNESS FOR LLM IN LONG-TERM TRAFFIC SIMULATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Long-term traffic simulation with large language model (LLM) agents requires stable routines and adaptation to disruptions under limited personal information. Existing agents use memory, planning, and reflection to sustain behavior, but retaining experience does not determine when acquiring information or invoking an LLM will improve an action. We propose an adaptive cognitive harness that couples persistent beliefs and habits with execution-supervised allocation of cognitive effort. Beliefs encode uncertain outcome expectations and habits encode context-dependent repetition; together with accessible evidence and constraints, they guide fast action, querying, LLM comparison, or their combination. We use paired simulator rollouts to train a lightweight predictor that estimates the task gains and additional resource costs of querying and LLM-based action comparison relative to a structured fast policy. During deployment, the predictor remains frozen while observed outcomes revise personal state. Across five 30-day world seeds evaluated in a paired design, the hybrid policy achieves lower mean task loss than all four mandatory-LLM controls while using approximately 40–44% fewer online LLM tokens, with quality gains concentrated in the disruption period. These results establish a favorable quality–cost trade-off for selective deliberation and support an execution-grounded approach to long-term LLM simulation: cognitive effort is allocated according to decision demands rather than uniformly spent on every action. Code will be released at https://anonymous.4open.science/r/Adaptive-Cognitive-Harness-CB35/.

1 INTRODUCTION

Long-term traffic simulation with LLM agents requires stable routines and adaptation under limited observations. One delay need not prompt a commuter to reconsider a route every morning, whereas repeated delays near a deadline may justify checking traffic or comparing alternatives. The challenge is to allocate information and computation where they improve behavior while preserving useful routines elsewhere.

Memory, planning, and reflection support continuity (Park et al., 2023; Piao et al., 2026), while mobility frameworks ground decisions in repeated route choices and urban execution (Wang et al., 2025; Liu et al., 2026). Yet retaining history does not determine when it is enough to act: routine repetition can miss a disruption, while repeated reconsideration may add no value. The unresolved issue is how experience should guide additional effort.

This allocation problem separates information acquisition from LLM comparison (Figure 1). A query may reveal a delay that a fast policy can handle; comparison may resolve a trade-off without new observations. Neither is automatically beneficial: reports can be stale, and an LLM cannot reveal missing evidence. Personal experience and deadlines determine when a familiar action merits either operation.

We propose an adaptive cognitive harness (Figure 2) that learns when to query and invoke an LLM. Beliefs encode uncertain outcomes and habits encode recurring actions. A fast policy reuses or locally adjusts an action; the allocator chooses FAST, query then FAST, LLM comparison, or query then comparison. Inputs comprise personal estimates, received reports, constraints, and a FAST

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

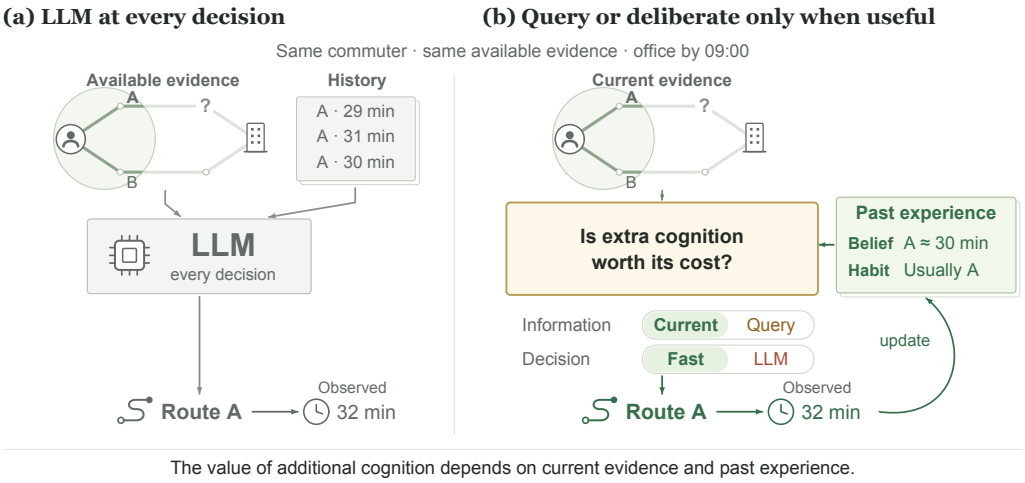

Figure 1: **Motivation.** The same commuter can select the same route with different cognitive effort. Our harness uses current evidence and persistent beliefs and habits to decide whether to query or invoke the LLM, and updates personal state after execution. The highlighted case uses current evidence and FAST; querying and LLM comparison remain separate choices, and all times are illustrative.

proposal—not hidden world conditions or others' private choices. Permissions and budgets restrict the available operations.

Training pairs permitted operations from the same simulator snapshot, with matched query evidence and a shared fixed-window continuation. A small predictor learns FAST-relative task gains and resource costs from execution, not LLM self-assessment. The reported evaluation prioritizes task gain and uses cost only to break ties. During deployment, the allocator and LLM remain fixed while personal outcomes update beliefs and habits, changing future actions and allocation inputs.

We evaluate task quality, disruption trajectories, and online resource use. FAST can change routes and LLM comparison can retain them; neither action changes nor fewer calls alone establish the value of personal history. Isolating that value beyond current reports and simple rules requires further controls.

The main contributions of this paper are as follows:

1. We specify an application-level integration of persistent personal state with a four-operation information-and-comparison space and explicit execution contracts.
2. We construct FAST-relative single-intervention supervision for a frozen gain/cost predictor, with personal state updated independently during deployment.
3. We report a bounded five-world quality–token comparison against mandatory LLM use. The incremental value of learning the gate or using personal history remains unisolated.

2 Related Work

Social simulation and persistent language agents. Generative Agents combines memory retrieval, reflection, and planning to support believable interactions in a small virtual community (Park et al., 2023). S^3 models emotion, attitude, and interaction in social networks, including population-level information propagation (Gao et al., 2023); AgentSociety expands the scale and range of LLM-driven social simulation (Piao et al., 2026). SOTOPIA instead evaluates goal-directed social interactions and illustrates why plausible dialogue alone is not sufficient evidence of social competence (Zhou et al., 2024). These efforts establish important behavioral and evaluation capabilities, but their respective aims do not settle when a persistent individual should pay for another information request or LLM comparison. CoALA organizes language agents around memory, action spaces, and decision processes (Sumers et al., 2024), providing an architectural perspective rather than an execution-trained allocator for this traffic setting. Habit theory distinguishes contextual repetition

from outcome evaluation (Wood & Rünger, 2016), and TRIPLE integrates planned and habitual behavior for personalization (Noh et al., 2026). We build on this distinction: beliefs and habits supply state for both action and effort allocation, without claiming that a structured habit variable or a modular agent architecture is itself new.

Traffic, mobility, and repeated route choice. SUMO provides microscopic, multimodal traffic execution (Lopez et al., 2018), while MATSim supports large-scale agent-based transport simulation with iterative plan execution, scoring, and replanning (Horni et al., 2016). Such systems make network constraints and repeated adaptation explicit; they do not by themselves specify how to allocate paid language-model cognition. LLMTraveler studies agentic LLM decisions in day-to-day route choice, and GATSim connects generative agents to urban mobility execution (Wang et al., 2025; Liu et al., 2026). Our work extends this line by asking when personal experience supports inexpensive action and when an information query, an LLM comparison, or both improve execution. Our contribution is not simply to add memory to traffic agents: we learn operation gains relative to an explicit same-information fast policy and retain separate query and comparator choices. Qi's laboratory studies motivate route persistence and attraction as behavioral questions (Qi et al., 2023; 2024); their individual records are not ground truth for our synthetic evaluation. Different networks, feedback protocols, and information permissions prevent treating published simulator-level scores as a common SOTA ranking, motivating matched controls within one environment.

Adaptive computation and execution-based feedback. Resource-rational analysis and rational LLM metareasoning formalize the trade-off between cognitive benefit and computational cost (Lieder & Griffiths, 2020; De Sabbata et al., 2024). FrugalGPT and RouteLLM economize inference through cascades or model routing (Chen et al., 2024; Ong et al., 2025); DPT-Agent uses fast and slower processes for real-time human–AI collaboration (Zhang et al., 2025). Our question differs from selecting a model or merely having two speeds: with one backbone fixed, which combination of information acquisition and action comparison improves a persistent agent's outcome? The supervision follows rollout-based policy learning: take an action and then follow a reference policy to estimate its consequences (Lagoudakis & Parr, 2003). Here, the interventions are bounded cognitive operations, and predictors receive only agent-accessible state. The allocator is trained offline and held fixed during deployment, while execution feedback continually updates the personal state that guides subsequent allocation decisions. ReAct interleaves reasoning and action, while Reflexion and Self-Refine use feedback to revise behavior or outputs (Yao et al., 2023; Shinn et al., 2023; Madaan et al., 2023). We instead use attributable execution outcomes for numerical state updates and gain/cost targets. The contribution is this specific coupling and its controlled evaluation, not the invention of metareasoning, memory, or feedback.

3 METHODOLOGY

3.1 PERSONAL INFORMATION AND PERSISTENT STATE

Let W_t denote the hidden world and $\mathcal{F}_t^i$ the information accessible to agent i. Public topology, personal executions, and delivered permitted reports are available; unobserved conditions and other agents' private choices are not. We form

$$z_t^i = \Phi(B_t^i, H_t^i, O_{\le t}^i, g_t^i, k_t^i), \qquad x_t^i = \psi(z_t^i, p_{0,t}^i), \tag{1}$$

where B, H are beliefs and habits, O contains accessible observations, g specifies goals, and k describes task progress. The unexecuted FAST proposal $p_{0,t}^i = \pi_{\mathsf{F}}(z_t^i)$ uses the same personal view; the deterministic extractor ψ does not access hidden world state W_t. Both z and x are lossy summaries, not sufficient statistics. Candidate actions $\mathcal{A}_t^i$ are constructed from permitted structure and subjective estimates; perceived feasibility need not reveal a hidden closure.

Beliefs are indexed by environment, origin/destination, mode, day type, and departure period. For an initialized duration estimate and a new personal outcome y,

$$\delta = y - \mu, \qquad \mu' = \mu + \alpha\delta, \qquad v' = (1-\alpha)(v + \alpha\delta^2). \tag{2}$$

Support, freshness, outcome variation, and epistemic uncertainty remain distinct. Delivered reports overlay relevant component estimates without becoming personal executions. A corresponding habit updates after eligible autonomous execution a_t in context c:

$$H'(c,a) = (1-\beta)H(c,a) + \beta\mathbf{1}[a = a_t]. \tag{3}$$

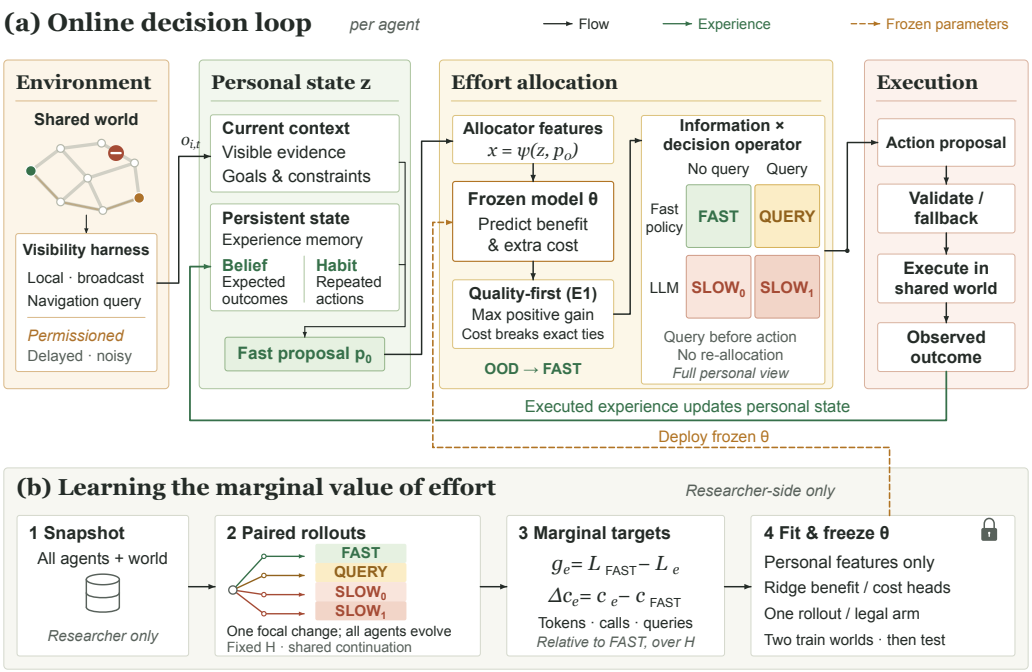

Figure 2: The implemented harness. E1 maximizes positive predicted task gain; cost breaks exact gain ties. Personal state changes, model parameters do not. Offline, one rollout per legal arm supplies FAST-relative supervision; fitting uses two worlds before a separate test. $\text{SLOW}_{0/1}$ and H denote $\mathsf{S}_0/\mathsf{S}_1$ and horizon h.

The pilot uses $\alpha = \beta = 0.2$. A surprising delay can revise belief while leaving habit intact; subsequent choices reshape habit. Forced detours and engine repairs have separate attribution. Appendix A gives the component forecasts and uncertainty heuristic.

3.2 FAST POLICY AND FOUR OPERATIONS

For a perceived-valid candidate a, let m_a be expected remaining minutes, $\sigma_{P,a}$ its predictive standard deviation, D time until the deadline, and η_i reliability aversion. Define

$$J(a) = m_a + \eta_i \sigma_{P,a}, \qquad R(a) = \text{clip}((m_a + \sigma_{P,a} - D)/15, 0, 1)\,. \tag{4}$$

With no deadline, $R = 0$; η_i defaults to 0.15. FAST first reuses a sufficiently supported habit when conflict and deadline risk are low and its score is within four minutes of the minimum. Otherwise it continues a suitable committed route within the same tolerance, or selects minimum J. If no perceived-valid candidate exists, FAST returns `WAIT_AND_RECHECK`. Appendix B specifies every condition and tie rule. Thus FAST can change actions and an LLM can rationally keep one.

The four operations are F (fast/no query), Q (fast/query), S_0 (LLM/no query), and S_1 (LLM/query). A quota ledger restricts legality but remaining quota is not a predictor feature. Both query arms target the lexicographically first eligible candidate route and request directed travel-time reports; candidate identities remain fixed. Accepted paired groups require identical returned-query and post-query-view hashes in Q/S_1, not merely the same channel distribution. Appendix C explains delayed delivery and keyed reconstruction.

The neutral prompt supplies candidates, personal estimates, constraints, and bounded evidence. The fixed Qwen3.5-35B-A3B comparator returns a selected identifier and structured explanation; invalid actions fall back to refreshed FAST. Internal thinking is disabled. Accordingly, "deliberation" means using this external comparator, not varying hidden reasoning depth. Provider latency does not advance simulated time; channel delivery delay does. Wall-clock latency is measured separately from travel or lateness (Appendix D).

3.3 EXECUTION-SUPERVISED OPERATION VALUE

For horizon h and fixed continuation π_c, branch e intervenes on the current focal decision only. The pilot's cost-sensitive value target is

$$V_h^{\pi_c}(z, e) = \mathbb{E}\Big[L_h(\mathsf{F}) - L_h(e) - \boldsymbol{\lambda}^\top \{C_h(e) - C_h(\mathsf{F})\} \mid z\Big]. \tag{5}$$

Here C_h counts the focal agent's tokens, queries, and provider attempts.

At a prospectively selected snapshot, legal branches share the initial world and personal view; other agents continue executing. Labels are $Y_{j,e} = (\Delta L_{j,e}, \Delta C_{j,e})$, where $\Delta L = L_h(\mathsf{F}) - L_h(e)$ and $\Delta C = C_h(e) - C_h(\mathsf{F})$. Cloud-output randomness is not controlled by restoring the simulator seed. The pilot has one rollout per legal arm, so a sample winner is not a known statewise optimum.

For decision time t_0, deadline d, and s the earlier of task resolution or horizon, the implemented focal loss is

$$L_h = (s - t_0) + 2\{[s - d]_+ - [t_0 - d]_+\} + 2N_{\text{switch}} + 120\,\mathbf{1}[\text{abort}]. \tag{6}$$

The lateness term is zero without a deadline; switches are autonomous. Incomplete tasks retain observed time and lateness, with zero additional censor penalty. Costs span the entire window, including any post-resolution cognition. This fixed-window cost convention captures downstream resource use but omits corresponding later-task benefit: it is a mixed-scope proxy, not agent welfare. Appendix F defines sign/ranking/choice sensitivity to this mismatch and distinguishes deployment's all-task objective.

Prediction and choice. Full has 20 features: seven current summaries, eleven belief-derived quantities, habit strength, and the FAST-to-minimum score gap. With train-only standardized inputs plus an intercept, each non-FAST operation has four ridge outputs:

$$\widehat{\Theta}_e = \arg\min_{\Theta_e} \sum_{j: e \in \mathcal{E}_j} \omega_j \|Y_{j,e} - \widetilde{x}_j^\top \Theta_e\|_2^2 + \rho \|\Theta_{e,\neg\text{bias}}\|_F^2. \tag{7}$$

The development protocol fixes $\rho = 10$ without validation. Write $G_e = \widehat{\Delta L}_e(x_t)$, $K_e = \boldsymbol{\lambda}^\top \widehat{\Delta C}_e(x_t)$, and $\mathcal{P}_t = \{e \in \mathcal{E}(z_t, b_t) \setminus \{\mathsf{F}\} : G_e > 0\}$. The quality-first selector used in E1 is

$$e_t^* = \begin{cases} \mathsf{F}, & \mathcal{P}_t = \varnothing, \\ \text{lex arg max}_{e \in \mathcal{P}_t}(G_e, -K_e, e), & \text{otherwise.} \end{cases} \tag{8}$$

The tuple prioritizes gain, then lower cost, then the fixed tie-break order $\mathsf{S}_0 \succ \mathsf{S}_1 \succ \mathsf{Q}$. Thus cost never outweighs a strictly larger gain. The pilot value selector maximizes $G_e - K_e$, with FAST valued at zero and favored on ties (Appendix E); it uses the same frozen heads. E1 has no recorded positive-gain ties, so its cost heads never resolve a choice (Appendix L). Unsupported categories or a standardized feature magnitude above eight cause a recorded FAST fallback. Acquisition precedes the chosen operator, not a second gate.

At fixed prices, these linear gain/cost heads are equivalent to ridge on net-value labels; the decomposition supports interpretation and price changes, not greater expressivity (Appendix E).

3.4 COMPUTATIONAL COMPLEXITY

Let K be the candidate count, $p = d + 1$ the feature dimension including the intercept, M the number of operations, and r the output count per non-FAST operation. Once the personal view is built, FAST and feature anchoring sort candidates in $O(K \log K)$ time; the ridge heads require $O((M-1)pr)$ time and parameter storage. For $d = 20$, $M = r = 4$, this is 252 coefficients, with no online gradient updates. The complete decision cost is

$$\begin{aligned} T_{\text{decision}} = T_{\text{view}} &+ O(K \log K + (M-1)pr) \\ &+ \mathbf{1}_{\text{query}} T_{\text{query}} + \sum_{j=1}^{n_{\text{LLM}}} T_{\text{LLM}}^{(j)} + T_{\text{verify}}. \end{aligned} \tag{9}$$

Here T_{view} includes candidate construction, history/report processing, and any post-query refresh. These costs are not constant in map or history size. LLM service time includes variable inference and provider waiting; token counts are a resource measure, not a wall-time complexity bound. Appendix I derives offline fitting and branch-acquisition costs and explains the implementation's history-dependent memory overhead.

4 EXPERIMENTS

We ask whether selective querying and LLM comparison can achieve good task outcomes without invoking the comparator at every decision, how performance changes during disruption and recovery, and what online resources this requires. Task quality is the primary endpoint; resource use is measured separately. We compare six conditions: hybrid, four mandatory-LLM controls, and the exploratory only FAST reference, which uses the structured FAST route-selection policy without active querying or route-level LLM comparison. Further mechanism-identification protocols are separated in Appendix G.

The preceding single-intervention pilot is reported separately in Appendix E.3: its fixed-price net gains are non-monotonic with training size, and mean value exceeds ridge at the two smaller sizes. One pilot test world does not establish a cross-world learning advantage. E1 below instead evaluates quality-first closed-loop policies.

4.1 ENVIRONMENT AND COMPARISON PROTOCOL

We evaluate in GATSim, which supports minute-level urban execution, persistent personal histories, and permissioned information channels (Liu et al., 2026). Each run contains ten fictional residents over 30 days, divided equally into stable, disturbance, and recovery phases. During disturbance, selected directed morning-route links incur $12 + \text{Uniform}(-4, 4)$ minutes of delay. Delays end on day 21 without resetting personal state. Five paired seeds (21001–21005) share task schedules and exogenous events, with 600 tasks per world.

The backbone is Qwen3.5-35B-A3B with internal thinking disabled. All policies retain public topology, completed-trip learning, and daily LLM reflection. Hybrid and the four mandatory-LLM controls share local, broadcast, and navigation access without injected noise, loss, or delay. Each route decision permits at most one active query and two provider attempts.

Policies. *hybrid* uses a frozen marginal-ridge predictor trained on 376 snapshots with quality-first selection. *slow + adaptive query* retains learned query selection on its own evolving state but forces LLM comparison. *slow + LLM query* delegates query selection to the LLM. *slow + always query* queries whenever legal, whereas *only slow* retains passive reports but never actively queries. The exploratory *only FAST* reference uses the structured FAST route-selection policy without active querying or route-level LLM comparator calls; personal learning and daily LLM reflection remain enabled.

Measurement. Task loss equals elapsed minutes plus twice lateness and a 120-minute terminal-failure penalty, excluding token and switching penalties. CVaR90 averages the largest 60 losses per world. Additional metrics include duration, morning lateness, and a post-hoc worst-resident excess-duration diagnostic (Appendix J.1). Online expenditure includes input/output tokens, reflection, and metered failures; queries are counted separately. Each policy contributes 3,000 tasks and 1,500 agent-days, with five paired worlds as the independent units. Training retains its legacy switching penalty, so the evaluation tests frozen-model transfer without refitting (Appendices E and F).

4.2 TASK PERFORMANCE AND DISRUPTION ADAPTATION

Table 1 shows that hybrid achieves the lowest observed mean loss of 39.088 minute-equivalents, compared with 39.155–39.764 for mandatory-LLM policies. Its reduction relative to slow + adaptive query is 0.587, with the same direction in all five paired worlds. Hybrid also obtains the lowest mean travel time (34.87 minutes) and morning lateness (4.21 minutes). These improvements show that selective deliberation preserves effective action selection: its lower loss reflects both shorter execution times and reduced lateness, rather than a trade-off between the two.

Table 1: Five-world task quality. Loss: mean ± SD; others: means. Loss/CVaR90 use minute-equivalents; others use minutes. Worst extra covers days 11–30 relative to days 1–10; other metrics cover all 30 days. Bold: lowest mean, not significance. †External reports excluded after day 10. Definitions and uncertainty: Appendix J.1.

Method	Loss ↓	Time ↓	AM late ↓	CVaR90 ↓	Worst extra ↓
slow + adaptive query	39.67 ± 0.42	35.07	4.61	84.64	5.345
slow + LLM query	39.15 ± 0.27	34.90	4.26	**81.95**	5.120
slow + always query	39.76 ± 0.36	35.10	4.66	85.54	5.455
only slow	39.43 ± 0.24	34.99	4.45	82.62	**4.830**
only FAST	43.84 ± 0.60	36.60	7.24	113.43	8.560
hybrid	**39.09 ± 0.34**	**34.87**	**4.21**	82.87	5.385

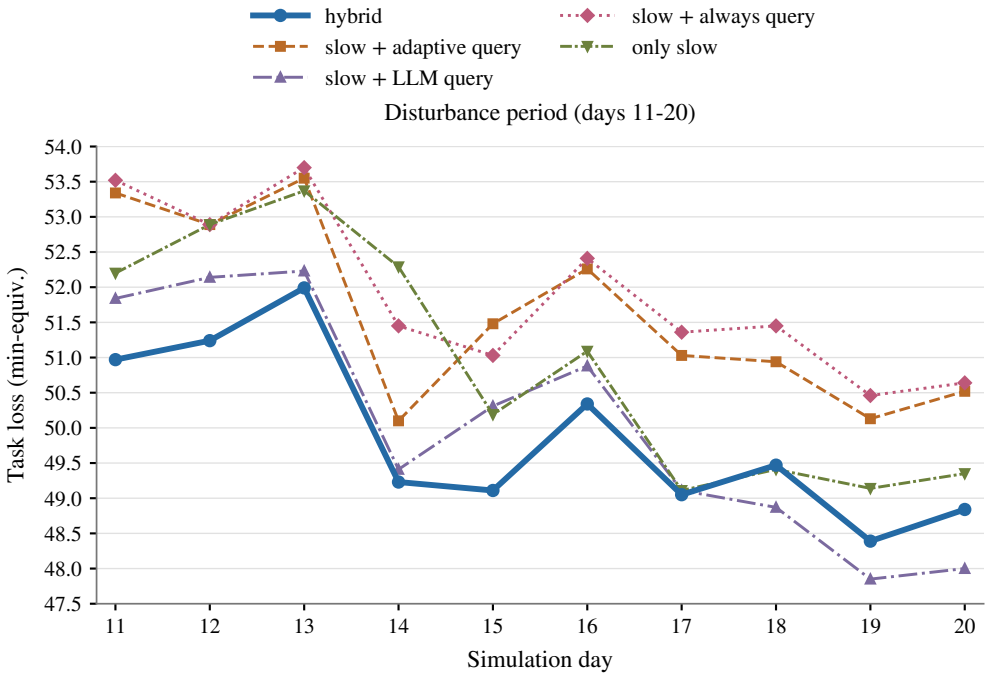

Figure 3: Unsmoothed daily mean loss over five worlds during disturbance (days 11–20): hybrid and four mandatory-LLM controls. Ticks: 0.5 minute-equivalents. Lines show means, not uncertainty; paired-world comparisons are reported in Appendix J.1.

The complementary metrics characterize different aspects of performance. Slow + LLM query attains the lowest CVaR90, whereas only slow achieves the lowest worst-extra score. Hybrid's distinctive strength is its joint lead on average loss, time, and lateness. For the mean-loss objective evaluated here, this combination makes selective invocation a competitive alternative to requiring LLM comparison at every decision.

Figure 3 locates this advantage within the disturbance period. Hybrid has lower daily mean loss than all four mandatory-LLM controls on days 11–17, sustaining its advantage across the early and middle disturbance period. Slow + LLM query attains lower daily mean loss than hybrid on days 18–20, indicating that the relative value of different configurations varies over time. The aggregate advantage is therefore supported by repeated favorable daily outcomes rather than a single isolated improvement. Together, these results show that selective deliberation can support effective execution under changing road conditions without making continuous LLM comparison a prerequisite.

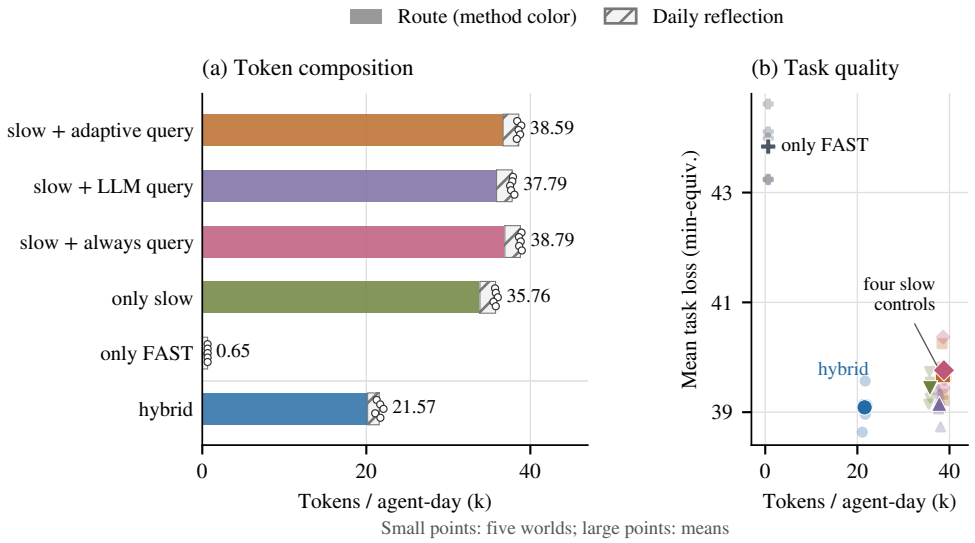

Figure 4: Online tokens per agent-day and mean task loss over five worlds. Bars split route/reflection use; small/large points show worlds/means. Tokens include input, output and metered failures. Only FAST has restricted information; total budgets are not equalized.

4.3 Measured online expenditure and operation use

Hybrid consumes 32,356,962 online tokens, or 21,571 per agent-day, reducing token expenditure by 39.69%–44.39% relative to the four mandatory-LLM policies. It averages 5.49 provider attempts per agent-day, compared with approximately 9.02–9.06 for these controls. Importantly, this reduction accompanies lower observed mean task loss. The resulting operating point combines lower observed mean task loss with substantially less online inference, illustrating the practical value of deciding when LLM comparison is worth invoking.

Figure 4 makes this joint quality–expenditure profile explicit. Hybrid's mean lies below and to the left of the four mandatory-LLM controls, while the token decomposition separates action-comparison expenditure from daily reflection. The only FAST reference occupies a much lower-token operating point, at about 649 tokens per agent-day, but has higher loss. This post-hoc reference excludes external reports after day 10, so its contrast with hybrid concerns the complete configurations, including information access. Reported expenditure covers online inference and excludes offline data collection and fitting.

The operation counts further explain what selective allocation changes. Hybrid issues 6.06 queries per agent-day, slightly more than the 5.90 of slow + adaptive query, despite substantially fewer provider attempts. Thus, computational savings coexist with continued information acquisition: the harness does not simply suppress all forms of cognitive effort. It treats querying and LLM comparison as separately allocable operations. The LLM-controlled query policy issues zero queries in these runs, so its realized behavior consists of LLM comparison without additional querying. Appendix J.3 provides detailed resource accounting, and Appendix J.4 reports operation allocation and resident-level differences.

4.4 Implications for cognitive allocation

E1 supports a practical conclusion: uniform LLM invocation is unnecessary for strong average execution performance in this setting. The hybrid policy combines lower observed mean loss with substantially reduced online token use. Because task loss excludes computational expenditure, this favorable combination is not built into the scoring rule by penalizing expensive reasoning.

A second implication is that information acquisition and action comparison should not be collapsed into a single measure of "thinking more." Hybrid can query more often than an adaptive-query

control while invoking the LLM less often, revealing a useful allocation pattern beyond a simple fast-versus-slow distinction. The evidence supports the joint harness as a favorable operating regime across five paired worlds with one backbone and one disruption family. The matched controls outlined in Appendix G provide the next step for separating the contributions of personal state, information access, and the allocation rule.

5 Conclusion

We introduced an execution-supervised harness coupling personal beliefs and habits with query and LLM-comparison allocation. Personal feedback changes state while the predictor remains frozen. Hybrid has the lowest observed mean loss and uses 39.69%–44.39% fewer online tokens than four mandatory-LLM controls. Primary quality effects are not significant, tails are not uniformly improved, and the only FAST contrast changes information access. This is an application-specific integration and bounded evaluation, not evidence that learning or personal history improves on same-information FAST, query-then-FAST, or simple gates. Those deployments remain necessary to identify the proposed mechanisms.

Future work. We will extend the evaluation to larger populations, longer horizons, and more diverse urban environments, examining how cognitive allocation scales with interacting agents and evolving conditions. Introducing social-information channels will enable controlled studies of how peer reports, source reliability, and information delays influence personal beliefs and deliberation. We will also develop finer-grained longitudinal metrics that distinguish routine persistence from unresponsive repetition and timely adaptation from unnecessary switching, alongside constraint satisfaction and computational expenditure. Further directions include budget-aware allocation, activity replanning, and evaluation against human route-choice data, broadening the framework's scope and strengthening its behavioral assessment.

AI use statement

Generative AI tools assisted research discussion, code development, drafting, translation, and formatting. The authors retain responsibility for the accuracy of the method descriptions, references, and all reported claims. Generative AI was not used to fabricate experimental outcomes.

Reproducibility statement

The supplement specifies state updates, FAST, the neutral comparator contract, training, and measurement. The source package includes aggregate E1 evidence and the frozen pilot summary. Complete runtime traces, baseline-specific prompts, channel settings, and source signatures are retained locally; a complete anonymous executable artifact has not yet been released. Aggregate reproducibility is not a full simulator reproduction.

Ethics statement

The protocol uses fictional residents and controlled simulation events, not identifiable human mobility records. Execution quality and adaptation in this setting do not validate predictions of real people or justify transport policy. Human-data extensions require distinct provenance, permissions, and matched evaluation.

References

Lingjiao Chen, Matei Zaharia, and James Zou. FrugalGPT: How to use large language models while reducing cost and improving performance. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=cSimKw5p6R.

C. Nicolò De Sabbata, Theodore R. Sumers, Badr AlKhamissi, Antoine Bosselut, and Thomas L. Griffiths. Rational metareasoning for large language models. *arXiv preprint arXiv:2410.05563*, 2024. URL https://arxiv.org/abs/2410.05563.

Chen Gao, Xiaochong Lan, Zhihong Lu, Jinzhu Mao, Jinghua Piao, Huandong Wang, Depeng Jin, and Yong Li. S^3: Social-network simulation system with large language model-empowered agents. *arXiv preprint arXiv:2307.14984*, 2023. URL https://arxiv.org/abs/2307.14984.

Andreas Horni, Kai Nagel, and Kay W. Axhausen (eds.). *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, 2016. doi: 10.5334/baw. URL https://doi.org/10.5334/baw.

Michail G. Lagoudakis and Ronald Parr. Reinforcement learning as classification: Leveraging modern classifiers. In *Proceedings of the Twentieth International Conference on Machine Learning*, 2003. URL https://cdn.aaai.org/ICML/2003/ICML03-057.pdf.

Falk Lieder and Thomas L. Griffiths. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:e1, 2020. doi: 10.1017/S0140525X1900061X. URL https://doi.org/10.1017/S0140525X1900061X.

Qi Liu, Can Li, and Wanjing Ma. GATSim: Urban mobility simulation with generative agents. *Transportation Research Part C: Emerging Technologies*, 186:105576, 2026. doi: 10.1016/j.trc.2026.105576. URL https://doi.org/10.1016/j.trc.2026.105576.

Pablo Alvarez Lopez, Evamarie Wiessner, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flotterod, Robert Hilbrich, Leonhard Lucken, Johannes Rummel, and Peter Wagner. Microscopic traffic simulation using SUMO. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2575–2582, 2018. doi: 10.1109/ITSC.2018.8569938. URL https://doi.org/10.1109/ITSC.2018.8569938.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. Self-refine: Iterative refinement with self-feedback. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2019. URL https://doi.org/10.52202/075280-2019.

Taehyung Noh, Seungwan Jin, Haein Yeo, and Kyungsik Han. TRIPLE: Theory-driven integration of planned and habitual behaviors for LLM-based personalization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(21):17626–17634, 2026. doi: 10.1609/aaai.v40i21.38818. URL https://doi.org/10.1609/aaai.v40i21.38818.

Isaac Ong, Amjad Almahairi, Vincent Wu, Wei-Lin Chiang, Tianhao Wu, Joseph E. Gonzalez, M. Waleed Kadous, and Ion Stoica. RouteLLM: Learning to route LLMs from preference data. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=8sSqNntaMr.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–22. Association for Computing Machinery, 2023. doi: 10.1145/3586183.3606763. URL https://doi.org/10.1145/3586183.3606763.

Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, Di Zhou, Chen Gao, Fengli Xu, Fang Zhang, Ke Rong, Jun Su, and Yong Li. AgentSociety: Large-scale simulation of LLM-driven generative agents advances understanding of human behaviors and society. *iFuture*, 2026. doi: 10.26599/IF.2026.9710004. URL https://doi.org/10.26599/IF.2026.9710004.

Hang Qi, Ning Jia, Xiaobo Qu, and Zhengbing He. Investigating day-to-day route choices based on multi-scenario laboratory experiments, part I: Route-dependent attraction and its modeling. *Transportation Research Part A: Policy and Practice*, 167:103553, 2023. doi: 10.1016/j.tra.2022.11.013. URL https://doi.org/10.1016/j.tra.2022.11.013.

Hang Qi, Ning Jia, Xiaobo Qu, and Zhengbing He. Investigating day-to-day route choices based on multi-scenario laboratory experiments, part II: Route-dependent attraction-based stochastic process model. *Communications in Transportation Research*, 4:100123, 2024. doi: 10.1016/j.commtr.2024.100123. URL https://doi.org/10.1016/j.commtr.2024.100123.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 8634–8652, 2023. doi: 10.52202/075280-0377. URL https://doi.org/10.52202/075280-0377.

Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L. Griffiths. Cognitive architectures for language agents. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=1i6ZCvflQJ.

Leizhen Wang, Peibo Duan, Zhengbing He, Cheng Lyu, Xin Chen, Nan Zheng, Li Yao, and Zhenliang Ma. Agentic large language models for day-to-day route choices. *Transportation Research Part C: Emerging Technologies*, 180:105307, 2025. doi: 10.1016/j.trc.2025.105307. URL https://doi.org/10.1016/j.trc.2025.105307.

Wendy Wood and Dennis Rünger. Psychology of habit. *Annual Review of Psychology*, 67: 289–314, 2016. doi: 10.1146/annurev-psych-122414-033417. URL https://doi.org/10.1146/annurev-psych-122414-033417.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=WE_vluYUL-X.

Shao Zhang, Xihuai Wang, Wenhao Zhang, Chaoran Li, Junru Song, Tingyu Li, Lin Qiu, Xuezhi Cao, Xunliang Cai, Wen Yao, Weinan Zhang, Xinbing Wang, and Ying Wen. Leveraging dual process theory in language agent framework for real-time simultaneous human-AI collaboration. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 4081–4108. Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.acl-long.206. URL https://aclanthology.org/2025.acl-long.206/.

Xuhui Zhou, Hao Zhu, Leena Mathur, Ruohong Zhang, Haofei Yu, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, and Maarten Sap. SOTOPIA: Interactive evaluation for social intelligence in language agents. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=mM7VurbA4r.