Explaining observed choices in behavioral logs using reinforcement learning–based computational models, and interpreting latent computational state transitions

GPT-5.2*

First Author†
First Affiliation
First Address
first@email

Second Author
Second Affiliation
Second Address
second@email

Third Author
Third Affiliation
Third Address
third@email

Fourth Author
Fourth Affiliation
Fourth Address
fourth@email

Abstract

Human decision-making behavior often reflects latent strategies that evolve over time rather than a single fixed policy. In this study, we investigate the automatic discovery of such latent decision-making strategies from behavioral log data using a switching reinforcement learning (RL) framework, and provide a computational neuroscience–inspired interpretation of the underlying mechanisms. We construct a sequential decision-making environment from large-scale behavioral logs and model user behavior as a mixture of multiple latent RL policies with dynamic switching. By estimating policy-specific parameters and their transition dynamics, the proposed approach identifies distinct decision strategies and their temporal shifts without explicit supervision. Furthermore, we interpret the learned parameters—such as reward sensitivity and prediction error dynamics—through the lens of computational neuroscience, linking observed behavioral patterns to theoretical constructs including reward prediction error and strategy arbitration. Experimental results on synthesized behavioral datasets demonstrate that the proposed framework effectively captures strategy heterogeneity and switching behavior, offering an interpretable account of decision-making processes beyond pure prediction. This work highlights the potential of switching RL as a principled tool for explaining complex human decision behaviors from real-world action logs..

1 Submission of papers to 2026 AI Co-Scientist Challenge Korea

Human behavior is shaped by multiple decision-making strategies that are flexibly deployed over time. Even within the same task environment, individuals alternate between exploratory and exploitative modes, reflecting dynamic internal states rather than random noise. However, most existing behavioral models assume a single stationary policy, limiting their ability to capture regime changes within individuals.

This study addresses this limitation by modeling learner behavior as switching among multiple latent decision-making strategies using a reinforcement learning–inspired framework. Rather than prespecifying strategy types or performing joint probabilistic inference, we adopt a data-driven approach that retrospectively identifies latent regimes from behavioral logs and analyzes their empirical transition

*Use footnote for providing further information, for less known open models (webpage, version)

†Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

Preprint. Under review.

dynamics. Furthermore, we interpret the discovered regimes through computational neuroscience concepts such as reward prediction error sensitivity and cognitive flexibility, providing mechanistic insight beyond prediction performance.

2 Related Work

2.1 Behavioral log analysis and decision models

2.1.1 Log-based behavioral modeling

Large-scale educational platforms generate rich sequential interaction logs, including correctness, response time, and content information. Knowledge tracing has been a dominant paradigm in this area, modeling how latent mastery evolves over time and enabling accurate prediction of learner performance. However, such approaches primarily focus on outcome prediction and offer limited insight into moment-to-moment decision strategies or behavioral regime changes within learners.

2.1.2 Markov Decision Process based approaches

To move beyond prediction, learner behavior has also been modeled from a sequential decision-making perspective, where observed actions are interpreted as trajectories generated by underlying policies. The Markov Decision Process framework provides a principled language for representing state, action, and reward, and motivates analyses that emphasize adaptation and interpretability rather than accuracy alone.

2.2 Reinforcement learning based models of human behavior

2.2.1 Q-learning and Rescorla–Wagner style updates

Reinforcement learning models such as the Rescorla–Wagner rule and Q-learning formalize trial-by-trial learning as value updating driven by reward prediction errors. These models have been widely used to interpret human behavior using cognitively meaningful parameters, including learning rate and reward sensitivity.

2.2.2 Softmax choice models

Softmax choice rules are commonly used to link latent values to observable actions, with inverse-temperature parameters capturing exploration–exploitation trade-offs. This parameterization supports interpretable behavioral claims while remaining scalable to large datasets.

2.2.3 Parameter-based cognitive interpretation

Despite their interpretability, standard reinforcement learning models typically assume a single stationary policy. This assumption limits their ability to capture regime changes frequently observed in human behavior, such as abrupt shifts in speed–accuracy trade-offs or exploration strategies.

2.3 Switching and mixture model families

2.3.1 Mixture of Experts

Mixture modeling provides a natural way to represent heterogeneous strategies by combining multiple "experts" under a gating mechanism that selects (or weighs) experts depending on context. The Mixture of Experts tradition formalizes how complex behaviors can be explained as compositions of simpler specialist policies, rather than a single monolithic strategy.

2.3.2 Hidden Markov Models

A complementary approach models switching directly through latent discrete states that evolve over time. Hidden Markov Models (HMMs) are a canonical tool for capturing regime persistence and structured transitions, and are widely used for sequential data where latent modes generate observable emissions (e.g., accuracy, response time, switching indicators).

2.3.3 Why switching RL and how we differ

Recent behavioral evidence suggests that human problem solving is often better described by switching among strategies rather than executing a stationary policy throughout. Many prior works follow a "assume a small set of strategies → fit/validate" pipeline, where candidate strategies are specified a priori and then tested against data. In contrast, our work emphasizes "automatic discovery of latent regimes + transition dynamics + behavioral interpretation" from real-world interaction logs, and then connects these within-learner regimes to between-learner heterogeneity captured by user-level clustering features.

3 Problem Formulation

3.1 Reinforcement Learning Environment Specification

We model learners' problem-solving behavior as a sequential decision-making process using interaction logs collected from an Intelligent Tutoring System. Each problem attempt is treated as a discrete decision step within a Markov Decision Process (MDP), characterized by a state s_t, an action a_t, and a reward r_t.

Because learners' internal cognitive states are not directly observable, the state is represented using observable behavioral and contextual features extracted from the logs, including recent correctness, response time, and problem metadata. The action corresponds to the learner's observed response choice, and the reward is defined as a binary performance signal indicating correctness.

All analyses are conducted in a logged interaction setting. The environment dynamics are treated as fixed but unknown, and the goal is not to learn an optimal policy through interaction, but to characterize decision-making behavior from observed trajectories.

3.2 Switching Reinforcement Learning Model

Learners' decision-making strategies are assumed to vary over time rather than remain stationary. To capture this non-stationarity, we adopt a switching reinforcement learning perspective in which observed behavior is generated by a sequence of latent strategies.

At each time step t, a learner's behavior is assumed to be generated by a latent strategy variable $z_t \in \{1, \ldots, K\}$, where each strategy corresponds to a distinct policy π_k. Based on preliminary clustering analysis, the number of strategies is fixed to $K = 4$ throughout the study. Latent strategies are inferred retrospectively from behavioral features, and each observed action is approximated as being generated by a single active strategy. Strategy switching is reflected by changes in z_t over time and is summarized via empirical transition frequencies $P(z_t \mid z_{t-1})$. This formulation enables analysis of both strategy heterogeneity and temporal switching without assuming a single stationary policy.

4 Methodology

4.1 Overview of the Switching RL Framework

The proposed framework identifies latent decision-making strategies from behavioral logs and analyzes how learners switch between them over time. The approach is fully data-driven and operates on observed interaction sequences, without requiring online learning or joint probabilistic inference. The analysis proceeds in three stages: (1) extraction of behavioral features from interaction logs, (2) assignment of latent strategies to each decision step, and (3) analysis of strategy-specific behavioral profiles and transition dynamics

4.2 Strategy-Specific Behavioral Characterization

Each latent strategy is characterized using the subset of interactions assigned to it. Rather than learning policies through reinforcement learning updates, we describe strategies by analyzing empirical distributions of behavioral signals within each strategy, including correctness, response time, and topic switching. Differences across strategies reflect distinct decision-making regimes, such as

stable high-effort behavior, fast low-deliberation responding, or unstable exploratory patterns. These empirical profiles provide interpretable representations of strategy-specific behavior grounded directly in observed data.

4.3 Strategy Transition Analysis

Temporal switching behavior is analyzed by examining transitions between latent strategies across consecutive decision steps. For each pair of strategies (i, j), we compute empirical transition frequencies $P(z_t = j \mid z_{t-1} = i)$ from the decoded strategy sequences.

These transition statistics capture key properties of decision dynamics, including strategy persistence, frequent switching, and asymmetric transitions between specific strategy pairs. No explicit transition model is assumed; all reported statistics are derived directly from the observed strategy sequences.

4.4 Interpretation of Effective Learning Parameters

Although reinforcement learning parameters are not explicitly estimated, strategy-specific behavioral profiles permit qualitative interpretation in terms of effective learning parameters. Strategies exhibiting strong post-error behavioral modulation and high switching probability following incorrect responses are interpreted as having high effective learning rates, reflecting strong sensitivity to feedback. Strategies characterized by consistent performance and low variability correspond to high effective reward sensitivity, indicating exploitative decision-making. This qualitative mapping allows latent strategies to be interpreted within a reinforcement learning framework while maintaining transparency and alignment with the observational nature of the data.

4.5 Summary of the Methodological Scope

The proposed methodology focuses on descriptive discovery and interpretation of latent decision-making strategies and their temporal dynamics. All components of the framework correspond directly to operations performed on logged interaction data, ensuring consistency between methodological description and implementation.

5 Computational Neuroscience Interpretation

The proposed framework yields, for each learner, a sequence of latent strategy states together with state-specific emission statistics and transition dynamics. These outputs permit mechanistic interpretation of behavioral logs within established computational neuroscience and reinforcement learning frameworks.

5.1 Reward Prediction Error and Value Updating

Learning in reinforcement learning is driven by the reward prediction error (RPE),

$$\delta_t = r_t - Q(s_t, a_t),$$

which quantifies the discrepancy between expected and obtained outcomes and is closely associated with dopaminergic signaling.

Although internal value functions are not directly observable in behavioral logs, state-wise emission statistics and transition patterns produced by the model provide behavioral proxies for RPE-related processes. Latent states differ systematically in mean reward emission. States with higher mean reward correspond to regimes in which internal value estimates are well aligned with environmental contingencies, whereas states with lower mean reward reflect persistent mismatches between expectation and outcome.

Latent states further differ in post-error behavioral modulation. Certain states exhibit increased response times following incorrect trials, whereas others show minimal post-error adjustment. Post-error slowing is a canonical marker of cognitive control engagement following negative outcomes and is commonly interpreted as reflecting large-magnitude negative RPE.

In addition, the probability of transitioning to a different latent state following an incorrect trial varies across states. Elevated post-error switching probability indicates that negative feedback

triggers reconfiguration of the current policy, whereas low post-error switching probability suggests maintenance of the current policy despite negative outcomes.

Together, these findings indicate that latent states are distinguished by their effective sensitivity to RPE, corresponding to distinct regimes of value updating and feedback utilization.

5.2 Strategy Switching and Cognitive Flexibility

The inferred latent-state sequences provide a direct characterization of temporal strategy switching. Switching behavior is quantified using three complementary measures: switch rate, defined as the probability that consecutive states differ; mean dwell time within a state; and entropy of the state occupancy distribution. Low switch rates combined with long dwell times indicate persistent engagement with a single strategy, consistent with habitual or automatized control. In contrast, high switch rates and short dwell times indicate frequent policy re-evaluation, consistent with flexible, goal-directed control. Switching behavior exhibits a non-monotonic relationship with performance. Learners with extremely high switch rates tend to occupy states associated with low mean reward emission and high tag-switch emission, suggesting unstable exploration or poorly structured search. Conversely, learners with low switch rates and high mean reward emission remain stably within high-performing states, indicating effective exploitation of a successful strategy. These results support a stability–flexibility trade-off in human learning. From a neural perspective, this trade-off is consistent with interactions between fronto-striatal circuits supporting flexible updating and posterior striatal circuits supporting habitual control. Strategy switching in the proposed framework can therefore be interpreted as a behavioral manifestation of dynamic weighting between these systems.

5.3 Effective Reinforcement Learning Parameters

Although explicit reinforcement learning parameters are not directly estimated, the statistical profiles of latent states permit qualitative mapping to effective reinforcement learning parameter regimes. States characterized by high mean reward emission and low tag-switch emission correspond to high inverse-temperature (β) regimes, in which action selection is strongly guided by learned values (exploitation). In contrast, states characterized by lower mean reward emission and higher tag-switch emission correspond to low-β regimes, reflecting noisier or more exploratory choice behavior. Latent states also differ in post-error behavioral modulation and post-error state-transition probability. States exhibiting pronounced post-error slowing and elevated post-error switching probability are consistent with high effective learning rates (α), indicating strong incorporation of feedback into subsequent behavior. Conversely, states exhibiting weak post-error modulation are consistent with low effective learning rates or reduced sensitivity to prediction errors. Taken together, latent states can be interpreted as occupying distinct regions in an effective (α, β) parameter space. This interpretation provides a principled link between unsupervised latent-state discovery and canonical reinforcement learning parameterizations, enabling cognitive characterization of learners without explicit subject-level model fitting.

6 Experiments

This section evaluates whether latent decision-making strategies and their switching dynamics can be recovered from real-world behavioral logs. Using sequential interaction traces, we assess (i) whether a switching-state model can identify interpretable latent regimes that differ in accuracy, response latency, and topic switching, (ii) how learners transition between these regimes over time, and (iii) whether the discovered structures align with meaningful between-learner heterogeneity as captured by user-level clustering features. Together, these experiments aim to move beyond accuracy-only summaries and provide mechanistic evidence for time-varying decision processes in human problem solving.

6.1 Experimental setup

6.1.1 Dataset construction and preprocessing

We use sequential interaction logs from the KT1 component of EdNet (Riiid's Education Dataset), merged with the contents/questions dataset to obtain ground-truth correctness labels and item-level

metadata. EdNet consists of large-scale, time-stamped learner–system interaction records collected from a real-world AI tutoring service, making it suitable for analyzing sequential problem-solving behavior. From the merged dataset, we randomly sampled 500 learners for the main analyses to balance scale with tractability. For each learner, interaction records were ordered by timestamp to form a single sequential trajectory, which we treat as one episode. We additionally conducted pilot analyses on a larger subsample of 2,000 learners and confirmed that the qualitative patterns of state-specific behavioral signatures and dominant transition tendencies were consistent with those observed in the 500-learner setting. To support switching-state inference, we construct three observed behavioral signals. First, correctness is defined as a binary indicator of whether the learner's response matches the ground-truth answer. Second, response time is represented as

$$\log\left(1 + \frac{\texttt{elapsed\_time}}{1000}\right),$$

where `elapsed_time` is measured in milliseconds; this transformation reduces skewness and stabilizes variance. Third, tag switching is defined as a binary indicator that equals 1 if the conceptual tag of the current interaction differs from that of the immediately preceding interaction, and 0 otherwise. Together, these signals capture performance, deliberation speed, and topical stability at each decision step.

6.1.2 Analysis scope and data usage

The primary objective of this study is unsupervised discovery and interpretation of latent behavioral regimes and their switching dynamics, rather than supervised prediction or performance benchmarking. Accordingly, models are fit and interpreted using the full 500-learner sample. While a user-level train/validation/test split could be applied to prevent cross-user leakage if held-out predictive evaluation were required, the results reported here focus on descriptive latent structure and within-learner temporal dynamics.

6.1.3 Models and hyperparameters

HMM (within-learner switching). To model within-learner strategy switching, we fit a Hidden Markov Model (HMM) with $K = 4$ latent states. At each interaction time step t, the HMM infers a discrete latent regime $z_t \in \{0, 1, 2, 3\}$ evolving according to a first-order Markov process. Each latent regime generates the observed behavioral signals, including correctness, log response time, and tag switching.

Clustering (between-learner heterogeneity). To capture between-learner differences, we compute user-level summary features such as overall accuracy, median response time, fast-response rate, early-to-late performance improvement, longest error streak, tag-switching rate, and recent accuracy volatility. We apply K-means and Gaussian Mixture Models (GMM) across multiple values of K to obtain interpretable learner profiles.

6.1.4 Baselines

To contextualize the proposed latent-state analysis, we established clustering-based baselines using user-level behavioral summaries. Specifically, we compared K-means and Gaussian Mixture Models (GMM) across a range of cluster numbers $K \in \{3, 4, 5, 6\}$. Model selection was guided by two standard unsupervised validity indices: the silhouette score (higher values indicate better cluster separation) and the Davies–Bouldin index (lower values indicate better clustering quality).

Across all tested values of K, K-means consistently outperformed GMM on both indices. The best-performing configuration according to these criteria was K-means with $K = 6$, achieving a silhouette score of 0.301 and a Davies–Bouldin index of 1.114.

For interpretability and direct comparison with the HMM-based analysis, which employs $K = 4$ latent states, we additionally report clustering results with $K = 4$ in the main text. This aligned setting enables a more direct behavioral-type comparison between static clustering-based approaches and the proposed latent-state switching framework.

6.2 Results

6.2.1 Clustering-based behavioral types (baseline)

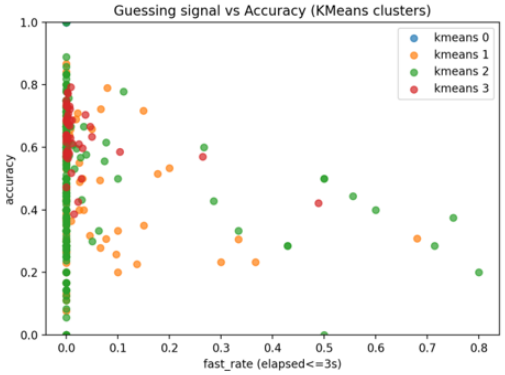

Figure 1: Clustering baseline: fast-response tendency versus accuracy by learner cluster ($K = 4$).

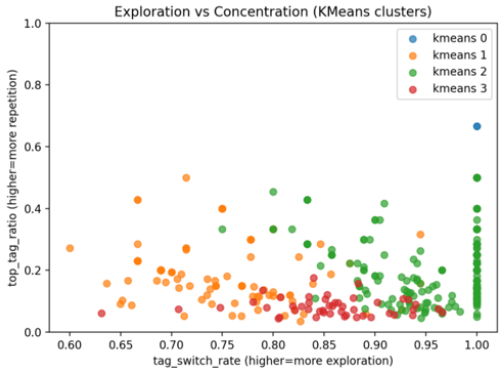

Figure 2: Clustering baseline: exploration versus repetition by learner cluster ($K = 4$).

As an aligned baseline, we report K-means clustering results with K=4 based on user-level behavioral summaries. Figure 1 shows the relationship between fast-response rate and accuracy across clusters, and Figure 2 illustrates exploration–concentration patterns via tag-switching rate and top-tag ratio. The clusters reflect relatively stable population-level profiles. The largest cluster exhibits higher accuracy and more concentrated topic engagement, while another cluster shows lower accuracy and higher volatility. A small cluster is characterized by extremely frequent tag switching with limited accuracy gains, indicating unstable exploratory behavior. Because clustering relies on aggregated summaries, it cannot capture within-learner temporal switching, motivating the HMM analysis below.

6.2.2 HMM-inferred latent states ($K = 4$)

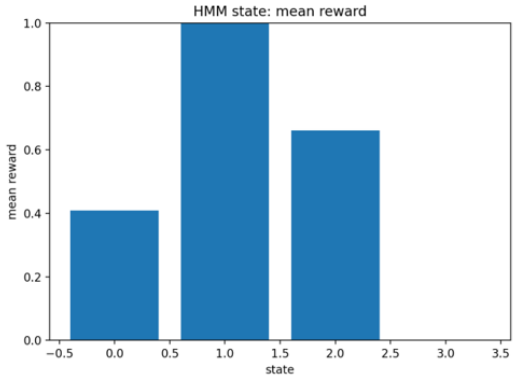

Figure 3: HMM state summaries: mean reward (or correctness) by inferred state ($K = 4$).

We next analyze latent decision regimes inferred by an HMM with K=4. Figure 3 summarizes the mean reward (correctness) associated with each inferred state. The inferred states differ clearly in performance. One state corresponds to a low-reward regime, another to an intermediate-reward regime, and one achieves near-perfect mean reward. Importantly, the high-reward state is not characterized by behavioral rigidity but co-occurs with frequent switching, indicating that high performance can coexist with flexible topic navigation. These results indicate that learner behavior is better described by switching among multiple latent

6.2.3 Transition dynamics

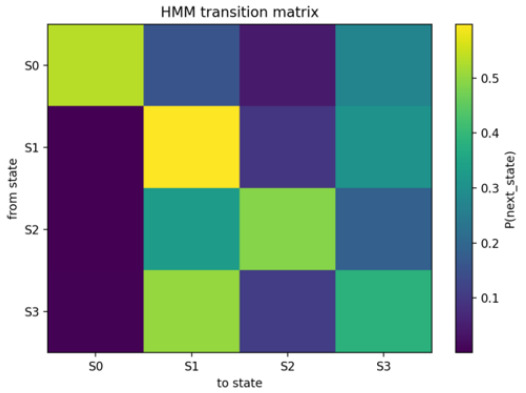

Figure 4: HMM transition matrix heatmap (rows: current state, columns: next state).

Figure 4 presents the HMM transition matrix, with rows denoting the current state and columns the next state. The transition matrix shows strong state persistence, with dominant diagonal entries. At the same time, structured off-diagonal transitions are observed, including prominent transitions from low-reward states to high-reward states. These patterns indicate systematic regime switching rather than random step-to-step noise.

6.2.4 What HMM adds beyond clustering

Compared to static clustering, the HMM captures within-learner regime changes that occur over sequential interactions. It reveals rapid transitions between stable, unstable, and fast-response regimes within the same learner, providing mechanistic evidence for time-varying decision strategies that cannot be recovered from user-level aggregates alone.

6.3 Qualitative Interpretation (Condensed)

6.3.1 Case selection protocol

To avoid cherry-picking, we adopt a simple and reproducible protocol. For each learner, we decode the most likely latent state sequence and identify (i) sustained dwell periods and (ii) salient switching events, focusing on recovery ($0 \rightarrow 3$) and collapse ($1 \rightarrow 0$) transitions. From these candidates, we select the earliest occurrence per learner and analyze fixed-length windows around each switch.

6.3.2 Representative switching patterns

Rather than interpreting latent states as fixed learner types, we examine short interaction windows to illustrate how regimes emerge and switch over time within individuals. Two recurrent patterns are observed: (i) Collapse-like switches ($1 \rightarrow 0$), where fast, low-deliberation responding is followed by low accuracy and increased instability, and (ii) Recovery-like switches ($0 \rightarrow 3$), where low-performance episodes transition into a high-accuracy regime.

6.3.3 Behavioral changes at the switch

Collapse transitions are typically associated with increased tag switching and longer response times without performance gains, indicating unstable exploration. In contrast, recovery transitions are marked by restored accuracy, while frequent topic switching often persists, suggesting that high performance does not require strict topical stability.

6.3.4 Summary

These qualitative patterns support the quantitative findings by showing that inferred states correspond to temporally coherent behavioral regimes, and that learning behavior is characterized by meaningful regime switches rather than stationary or purely noisy dynamics.

7 Conclusion

This work presents a switching latent-state framework for discovering and interpreting time-varying decision-making strategies from large-scale educational interaction logs. By modeling behavior as transitions among latent strategy states, the proposed approach moves beyond accuracy-centered prediction and provides a mechanistic account of how learners adapt over time. The results demonstrate that (i) multiple latent decision strategies can be automatically discovered from behavioral logs without prior specification, (ii) learners differ not only in the strategies they employ but also in their switching dynamics, and (iii) state-specific emission and transition patterns admit coherent interpretation in terms of reward prediction error sensitivity, cognitive flexibility, and effective reinforcement learning parameter regimes. These findings indicate that computationally meaningful cognitive variables can be recovered from behavioral data alone, without access to neural measurements. The proposed framework therefore establishes a scalable link between educational data mining and computational neuroscience. Future work will extend this approach by integrating neural or physiological measurements for direct validation of state-level interpretations, developing online inference procedures for real-time tracking of cognitive states, and constructing personalized models of strategy dynamics to support adaptive educational interventions.

References

[1] Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2015). Reinforcement learning in the brain. *The Journal of Neuroscience*, 35(34), 11511–11520. https://doi.org/10.1523/JNEUROSCI.2371-15.2015

[2] Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, 173, 34–42. https://doi.org/10.1016/j.cognition.2017.12.014

[3] Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, 143(6), 2074–2081. https://doi.org/10.1037/a0038199

[4] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press. http://incompleteideas.net/book/the-book-2nd.html

AI Co-Scientist Challenge Korea Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [N/A] .
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[N/A] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "AI Co-Scientist Challenge Korea paper checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the paper's main contributions: discovering time-varying latent decision regimes from behavioral logs and interpreting regime-specific patterns via reinforcement learning and computational neuroscience constructs (Abstract; Section 1). The claims are aligned with the descriptive nature of the method and are supported by the empirical results (Sections 7–8).

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [No]

   Justification: The work has limitations (e.g., a fixed number of latent states $K = 4$, reliance on observational logs and model assumptions, and evaluation on a single main dataset), but these are not currently discussed in a dedicated *Limitations* section. Adding an explicit limitations paragraph or section would strengthen transparency.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: The paper does not present formal theoretical results (e.g., theorems or lemmas) requiring proofs; the contributions are empirical and interpretive in nature (Sections 5–7).

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Question: Are the data and code openly accessible?

Answer: [No]

Justification: While the dataset used in the study is publicly available (Section 3.1), the analysis code and full reproduction scripts are not currently provided. The paper instead supports reproducibility through detailed methodological descriptions (Sections 5 and 7.1).

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While AI Co-Scientist Challenge Korea does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: While the dataset used in the study is publicly available (Section 3.1), the analysis code and full reproduction scripts are not currently provided. The paper instead supports reproducibility through detailed methodological descriptions (Sections 5 and 7.1).

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.

- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental setting specifies the dataset, subsampling protocol (e.g., 500 learners), observed behavioral signals (correctness, log response time, and tag switching), and model and hyperparameter choices (e.g., an HMM with $K = 4$) (Sections 7.1.1–7.1.3). Baseline configurations and selection criteria are also clearly described (Section 7.1.4).

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: The primary contribution of the paper is unsupervised regime discovery and descriptive interpretation rather than hypothesis testing or supervised performance benchmarking. As such, classical significance testing or error bars are not central to the main claims (Sections 7.2–7.3).

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: The paper does not currently specify hardware details (CPU/GPU), memory usage, or run-time required to reproduce the experiments. Including a brief paragraph describing the compute environment and approximate runtime for key experiments would address this item.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://nips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The study analyzes publicly released and anonymized educational interaction logs and does not involve interventions or deceptive procedures. The work is consistent with responsible research practices (Section 3; Appendix/Checklist statement).

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: The paper does not currently include a dedicated discussion of positive and negative societal impacts. Adding a brief *Broader Impacts* paragraph would clarify potential benefits (e.g., interpretability for learning analytics) and risks (e.g., misuse for high-stakes profiling).

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release high-risk generative models or sensitive scraped datasets and therefore does not require special safeguards beyond standard citation and data-use compliance (Sections 3 and 12).

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper uses the EdNet (KT1) dataset and cites the original dataset release, which is distributed under the Creative Commons Attribution–NonCommercial 4.0 International license. The dataset source is appropriately credited (Section 3.1; References).

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not introduce or release new datasets, models, or code assets; it reports analyses conducted on existing public data (Sections 3 and 7).

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The work does not conduct crowdsourcing or recruit human subjects; it relies exclusively on previously collected and publicly released interaction logs (Section 3.1).

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The study does not involve direct human subject experimentation and relies on an anonymized public dataset. Therefore, IRB approval applicable to prospective human-subject research is not required (Section 3.1).

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.