# PJDSC 2025

# Project Heat Resilience PH: A Reinforcement Learning Approach towards Heat Index Forecast's Adaptive Model Selection (HI'FAMS)


***PROXIMA LABS***

*Ducay, Harvey Lemuel Ogao*

*Empeño, Jhon Kheil Cymon Tomarong*

*Monfero, John Benedict Angeles*

## EXECUTIVE SUMMARY

The Philippines—particularly its densely populated capital, Manila—is acutely vulnerable to extreme heat events intensified by climate change. Under the Project Heat Resilience PH initiative, this study advances a **data-driven forecasting approach** designed to strengthen heat-risk preparedness. Rather than optimizing suspension policies directly, we focus on **improving next-day Heat Index (HI) forecasts,** recognizing that more accurate and reliable predictions are foundational to timely, proportionate decisions by education and public-health authorities. Static heat-index thresholds remain important for action, but their effectiveness depends on the **quality of the forecasts** that trigger them.

This work reframes the problem as **Adaptive Model Selection (AMS):** each day, the system selects **one best predictive model** from a portfolio (the "model zoo") based on the day's context (e.g., meteorological features, season/regime flags, and recent model errors). We implement this as a contextual bandit—using lightweight, data-efficient policies such as LinUCB and Contextual Thompson Sampling—that learns from daily outcomes without requiring a full simulation of real-world interventions. The learning signal is a **differential reward** that compares the chosen model's error against a **Best-in-Class Benchmark:**

$$R_{t+1} = \left(y_{t+1} - \hat{y}_{t+1}^{a^*}\right)^2 - \min_{m \in M} \left(y_{t+1} - \hat{y}_{t+1}^{m}\right)^2$$

so **positive rewards** indicate improvement over the benchmark. This formulation stabilizes learning, stays computationally tractable, and directly targets **predictive utility**. (A threshold-aware penalty near critical HI cutoffs—e.g., 42 °C and 46 °C—can be incorporated as an extension when stakeholders require it.)

Built on daily Open-Meteo inputs and historical observations for Metro Manila, the pipeline **runs all candidate models**, publishes only the selected forecast, and updates the policy when truth arrives the next day—enabling transparent **regret tracking** and steady improvement. In doing so, the system aims to **reduce forecast errors precisely when decisions matter most**, thereby improving the information available to decision-makers (e.g., schools, LGUs, and health agencies). While **policy enactment** (such as class suspensions) remains outside this project's operational scope, the proposed AMS framework provides a **robust, reproducible, and testable** foundation for **more effective, timely, and justified** heat-risk responses across the Philippines.



## KEYWORDS

Adaptive Model Selection (AMS), Contextual Bandits, Heat Index Forecasting, Metro Manila, Philippines, Extreme Heat Events, Climate Resilience, Decision Support Systems, Differential Reward, Benchmark Model (Best-in-Class), Regret Minimization, Time-Series Forecasting, Numerical Weather Prediction (NWP), LinUCB, Contextual Thompson Sampling (CTS), Threshold-Aware Penalties (42 °C, 46 °C), Open-Meteo Data, Model Zoo, Education Policy (Class Suspension), Public Health Preparedness, Project Heat Resilience PH

## BACKGROUND

The Republic of the Philippines is consistently ranked among the nations most vulnerable to climate change. In urban centers like Manila—the world's densest city—the **urban heat island** effect amplifies regional heatwaves, driving the Heat Index (HI) to dangerous levels that threaten public health, especially for children and adolescents in schools. In response, authorities from the Department of Education (DepEd) to Local Government Units (LGUs) have relied on class suspensions to mitigate heat-related risks.

However, the decision framework behind these suspensions is often **reactive**. Policies typically hinge on **static HI thresholds** (e.g., "Danger" and "Extreme Danger"), which provide clarity but lack dynamism: they do not account for the **expected duration**, **peak timing**, or **school-day context** (e.g., examinations, special schedules), nor do they explicitly incorporate **forecast uncertainty**. As a result, suspensions may be announced **too late**—after substantial exposure—or enacted **unnecessarily**, disrupting instruction and family routines without proportional health benefit. Crucially, the effectiveness of any threshold policy is only as strong as the **accuracy and reliability of the forecast** that triggers it.

**Project Heat Resilience PH** addresses this gap by shifting from control-oriented reinforcement learning in simulated environments to a **forecast-improvement approach** grounded in **Adaptive Model Selection (AMS)**. Rather than learning when to suspend classes, the system learns **which forecasting model to trust each day** for next-day HI. Formulated as a **contextual bandit** problem, the selector draws on daily meteorological context (e.g., Open-Meteo features, seasonal/regime indicators, recent model errors) to choose a single forecaster from a curated **model portfolio**. Learning is guided by a **differential reward** that directly measures improvement over a **Best-in-Class benchmark**, enabling stable, data-efficient updates without the complexity of full environment simulation. By **reducing forecast error precisely where it matters most**, AMS strengthens the informational foundation on which DepEd and LGUs make time-sensitive decisions—supporting more **timely, proportionate, and defensible** responses to extreme heat, while keeping policy enactment itself outside this project's operational scope.

## THE PROBLEM

**Real-world challenge:** Philippine education authorities face a dynamic trade-off during extreme heat:

- **Minimize student health risk** (heat illness during commute/class time).
- **Minimize educational disruption, false alarm for overprediction and lack of preparedness for underforecast Heat Indexes** (such as delayed or unnecessary suspensions, lost

instruction hours, family logistics, calendar slippage).

**Our research problem.** Rather than issuing suspension decisions directly, we address the **information bottleneck** that drives those decisions: **improving next-day Heat Index (HI) forecasts** via **Adaptive Model Selection (AMS)**. Bad forecasts—especially near critical thresholds (e.g., 42 °C, 46 °C)—cause costly miscalls (either late/insufficient action or unnecessary suspensions). Our problem is therefore:

**Pick one model each day** from a fixed portfolio to **minimize expected, threshold-aware forecast loss** (or, equivalently, **maximize differential improvement** over a Best-in-Class benchmark).

1. Let $x_{t-1}$ be needed, the day before the present day where $x_t$ holds a lists of context features be available in Open-Meteo
2. Let $M = \{m_1, m_2, \ldots, m_K\}$ holds all candidate models that Open-Mateo predicts the Heat Index given the $x_{t-1}$ features. Which yields forecasted Heat Index $\hat{y}^{m}_{t+1}$ while waiting the $y_{t+1}$ actual observed Heat Index
3. Through Reinforcement Learning's Proximal Policy Optimization, having a decision $a^* \in M$
4. The Goal, minimize the regret function $R_{t+1}$

# OBJECTIVES

The primary objectives of **Project Heat Resilience PH**—reflecting the scope shift from control-policy PPO to **Adaptive Model Selection (AMS)** for next-day Heat Index (HI) forecasting—are:

1. **Design an AMS pipeline for daily HI forecasting.** Formalize next-day HI prediction as a **contextual bandit** that selects one forecaster $a^*$ from a fixed **model portfolio** (model zoo) using day $t-1$ meteorological context and recent error signals.
2. **Engineer a differential reward aligned to predictive utility.** Implement the **benchmark-referenced reward** $R_{t+1} = (y_{t+1} - \hat{y}^{a^*}_{t+1})^2 - \min_{m \in M} (y_{t+1} - \hat{y}^{m}_{t+1})^2$ to directly incentivize improvement over a **Best-in-Class** model, with an optional **threshold-aware** extension for HI cutoffs (e.g., 42 °C, 46 °C).
3. **Develop and train lightweight selection policies.** Implement **LinUCB** and **Contextual Thompson Sampling (CTS)** for daily updating; include a **safe fallback** (baseline guard) and a deterministic **tie-breaker** (risk, variance, regime bias, switching cost).
4. **Build a reproducible, operational data stack.** Integrate **Open-Meteo** feature ingestion, model-zoo inference, and policy selection with **Supabase** (data/logging) and **Vercel** (scheduled jobs), ensuring versioning of **models**, **policy state**, and **published forecasts**.
5. **Evaluate against strong baselines with rigorous metrics.** Compare the AMS policy to (i) the fixed benchmark, (ii) best-historical per regime, and (iii)

median ensemble using MAE/MSE, daily & cumulative regret, and threshold-misclassification rates; apply walk-forward backtesting and, where applicable, Diebold–Mariano significance tests.

6. **Deliver an analytical dashboard for transparency and uptake.** Provide a functional dashboard that reports **selected forecasts**, **per-model error leaderboards**, **regret trends**, **threshold errors**, data latency, and job health—supporting stakeholder interpretation and adoption.
7. **Document scope, governance, and reproducibility.** Publish schemas, APIs, and evaluation protocols; clearly state that policy enactment (e.g., suspensions) remains with authorities, while the system's role is to **improve forecast quality** that informs those decisions.

## SCOPE & LIMITATIONS

**Scope:**

**> Geographical Focus:** Manila City

**> Forecasting Focus:** Next-day **Heat Index (HI)** point forecasting; inputs include Open-Meteo variables (e.g., temperature, relative humidity) and engineered regime/context features.

**> Decision Space: Adaptive Model Selection (AMS)**—the policy selects **one** forecaster per day from a predefined **model portfolio** (model zoo); no direct control of suspensions.

**> Data & Operations:** Daily batch pipeline (ingest → predict → select → publish; verify → learn next day) with logging/versioning; selected forecast is exposed via a public API for downstream use.

**Limitations:**

**> Finite Portfolio:** Results are bounded by the quality/diversity of the available forecasting models; the system does not train new forecasters end-to-end.

**> Benchmark Dependence:** The differential reward relies on a chosen **Best-in-Class** benchmark; if this benchmark is regime-biased, the learning signal may be skewed.

**> Data/Forecast Dependency:** Performance depends on Open-Meteo inputs and observation quality/availability; outages or representativeness issues can degrade accuracy.

**> Non-stationarity:** Abrupt regime shifts or structural climate changes may reduce performance; explicit domain adaptation is out of scope for this iteration.

**> Single-Hazard, Daily Cadence:** The system targets **heat index only** (no typhoons/flooding/air quality) and operates at **daily** resolution (no sub-daily or real-time probabilistic forecasts).

**> Governance Boundary:** The system **does not** automate class suspensions; final policy decisions remain with DepEd/LGUs and should consider multiple information sources.

## METHODOLOGY

This research utilizes a **model-selection Reinforcement Learning** methodology. The process keeps the original structure (Formulation → State → Action → Reward → Environment → Training → Prototype), but refocuses from suspension control to **Adaptive Model Selection (AMS)** for next-day Heat Index (HI) forecasting.

**1.1. Problem Formulation (MDP):** The forecasting task is cast as a **daily, single-step decision process**: on each day ttt, the agent observes context (the day's meteorological summaries and calendar position) and **chooses one**

**predictive model** to publish the next-day HI forecast. Although this is operationally a contextual bandit, we preserve the MDP framing used in the original draft by treating each day as a one-step episode: *(state → action → observe actual → reward → update).*

**1.2. State Space (S):** The state vector provides a comprehensive view of day-t conditions:

> **Calendar/Seasonality:** *Day of Year* in the range [0, 365] (cyclically encoded for seasonality); *Year* for slow drift.

> **Meteorological Summaries (day** t**):**

- Relative Humidity @2 m (**mean / min / max**),
- Apparent Temperature / Heat Index (**mean / min / max**),
- Temperature @2 m (**mean / min / max**),
- Rain (**mean / min / max**),
- Cloud Cover (**mean / min / max**),
- Wind Direction @100 m (**mean / min / max**, represented as degrees),
- Wind Speed @100 m (**mean / min / max**).

**1.3. Action Space (A):** The agent chooses from **all candidate forecasting models** available in the portfolio (the "model zoo") generated from the Open-Meteo–driven pipeline and served through Hugging Face Inference/Space.

**1.4. Reward Function (R): We retain the project's differential reward**, computed after the next day's observation becomes available:

$$R_{t+1} = (y_{t+1} - \hat{y}_{t+1}^{a*})^2 - \min_{m \in M} (y_{t+1} - \hat{y}_{t+1}^{m})^2$$

That is $\min_{m \in M} (y_{t+1} - \hat{y}_{t+1}^{m})^2$ is the MSE of the **Best-in-Class Benchmark** model and $(y_{t+1} - \hat{y}_{t+1}^{a*})^2$ is the MSE of the **chosen** model. A **positive** reward indicates the selected model outperformed the benchmark. *(No additional terms are introduced beyond this project-defined reward.)*

**1.5. Environment Implementation:** The MDP is instantiated as a daily model-selection environment:

**Ingest & Predict (12:00 PHT, current day)**

- Orchestrated by **Vercel Cron** calling `/api/forecast`.
- Fetch Open-Meteo features; validate and store in **Supabase** (`features_daily`).
- **Hugging Face** service generates next-day predictions for **all** candidate models; store in `model_predictions`.

**Select & Publish**

- The **RL model-selection policy** (bandit) receives the state and scores the candidates.
- The chosen model's forecast is **published** and exposed via the website/API; decision and context are logged in `policy_actions`.

**Verify & Learn (next day when actual $y_{t+1}$ is available)**

- Obtain actual HI; compute the project reward $R_{t+1}$ and per-model errors; write to `evaluation_daily`.

- **Update** policy parameters and version in `policy_state`; refresh metrics dashboards and alerts.

**2. Prototype Implementation and User Interface (UI):**

**> Frontend:** React **Next.js** + Tailwind on **Vercel**; a public route `/api/forecast` renders the selected forecast with model and policy versions.

**> Compute & Data: Vercel Cron** for the two jobs (Ingest & Predict; Verify & Learn), > **Hugging Face** for candidate model inference, **Supabase Postgres** for data and logs.

**> Dashboard:** leaderboards of per-model errors, the daily **benchmark-differential reward**, cumulative trends, job health, and latency—mirroring the architecture diagrams you provided.

## REFERENCES


***Agrawal, S., & Goyal, N. (2013). Thompson sampling for contextual bandits with linear payoffs. Proceedings of the 30th International Conference on Machine Learning (ICML). PMLR.***

***Agarwal, A., Luo, H., Neyshabur, B., & Schapire, R. E. (2017). Corralling a band of bandit algorithms. Proceedings of the 30th Conference on Learning Theory (COLT). PMLR.***

***Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (2002). The nonstochastic multiarmed bandit problem. SIAM Journal on Computing, 32(1), 48–77.***

***Ben Bouallègue, Z., Magnusson, L., Bremnes, J. B., et al. (2024). The rise of data-driven weather forecasting: A first statistical assessment of ML-based forecasts in an operational-like context. Bulletin of the American Meteorological Society, 105(6).***

***Bi, K., Chen, Y., Deng, X., et al. (2023). Accurate medium-range global weather forecasting with 3D neural networks. Nature.***

***de Burgh-Day, C. O., Pook, M. J., Jakob, C., & May, P. T. (2023). Machine learning for numerical weather and climate modelling. Geoscientific Model Development, 16, 6433–6477.***

***Dudík, M., Langford, J., & Li, L. (2014). Doubly robust policy evaluation and optimization. Statistical Science, 29(4), 485–511.***

***Foster, D. J., Krishnamurthy, A., & Luo, H. (2019). Model selection for contextual bandits. Advances in Neural Information Processing Systems (NeurIPS).***

***García, J., & Fernández, F. (2015). A comprehensive survey on safe reinforcement learning. Journal of Machine Learning Research, 16, 1437–1480.***

*Iyengar, G. N. (2005). Robust dynamic programming. Mathematics of Operations Research, 30(2), 257–280.*

*Jiang, N., & Li, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. Proceedings of the 33rd International Conference on Machine Learning (ICML). PMLR.*

*Kassraie, P., Emmenegger, N., Krause, A., & Pacchiano, A. (2023). Anytime model selection in linear bandits. Advances in Neural Information Processing Systems (NeurIPS).*

*Kumar, A., Zhou, A., Tucker, G., & Levine, S. (2020). Conservative Q-learning for offline reinforcement learning. Advances in Neural Information Processing Systems (NeurIPS).*

*Lam, R., Das, A., Nguyen, T., et al. (2023). Learning skillful medium-range global weather forecasting. Science. (GraphCast).*

*Laroche, R., Trichelair, P., & Tachet des Combes, R. (2019). Safe policy improvement with baseline bootstrapping (SPIBB). Proceedings of the 36th International Conference on Machine Learning (ICML). PMLR.*

*Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. arXiv preprint arXiv:1003.0146.*

*Nilim, A., & El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. Operations Research, 53(5), 780–798.*

*Pathak, J., Subramanian, S., Harrington, P., et al. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv preprint arXiv:2202.11214.*

*Raftery, A. E., Kárný, M., & Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics, 52(1), 52–66.*

*Timmermann, A. (2006). Forecast combinations. Handbook of Economic Forecasting, Vol. 1. Elsevier.*

*Watson-Parris, D. (2021). Machine learning for weather and climate are worlds apart? Philosophical Transactions of the Royal Society A, 379(2194), 20200098.*

*Open-Meteo. (n.d.). Weather forecast API documentation. (Accessed 2025).*

*Hugging Face. (n.d.). Inference Endpoints API reference. (Accessed 2025).*

*Supabase. (n.d.). Supabase Docs: Database & REST API references. (Accessed 2025).*

*Department of Education (DepEd), Philippines. (2022). DepEd Order No.*

*037, s. 2022 – Guidelines on the cancellation or suspension of classes and work in schools in the event of natural disasters, power outages/power interruptions, and other calamities.*

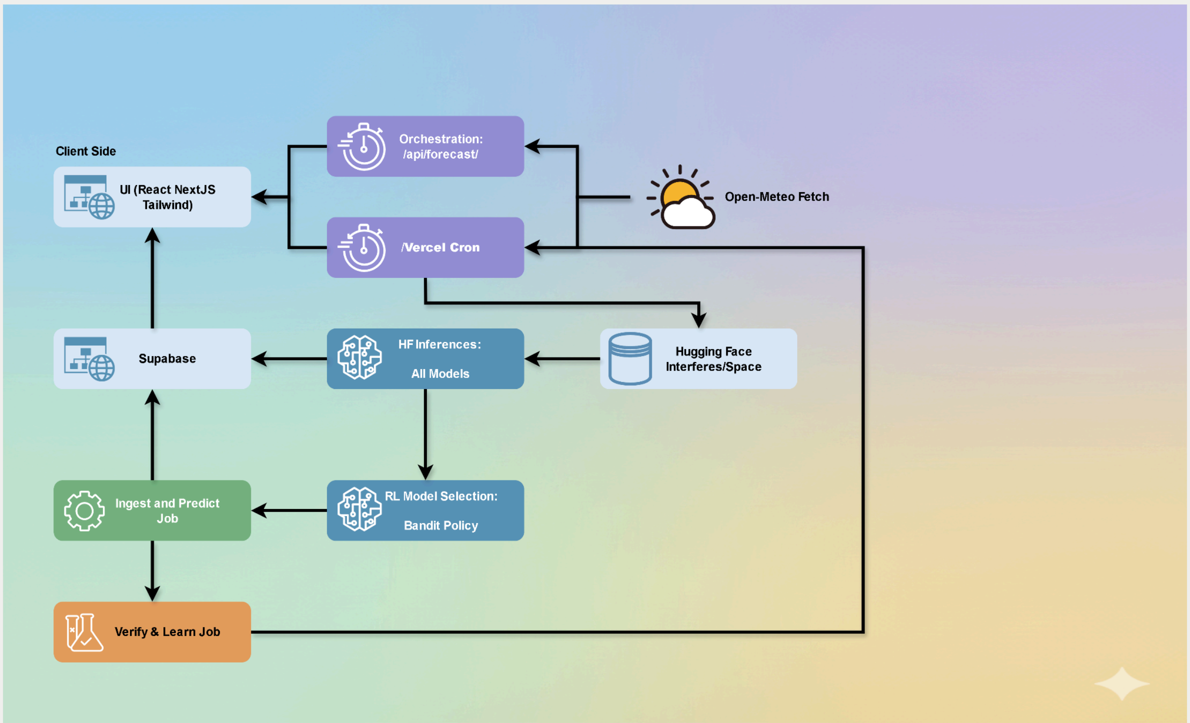
*Intergovernmental Panel on Climate Change (IPCC). (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report. Cambridge University Press.*

*Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA). (n.d.). Heat index information. (Accessed 2025).*

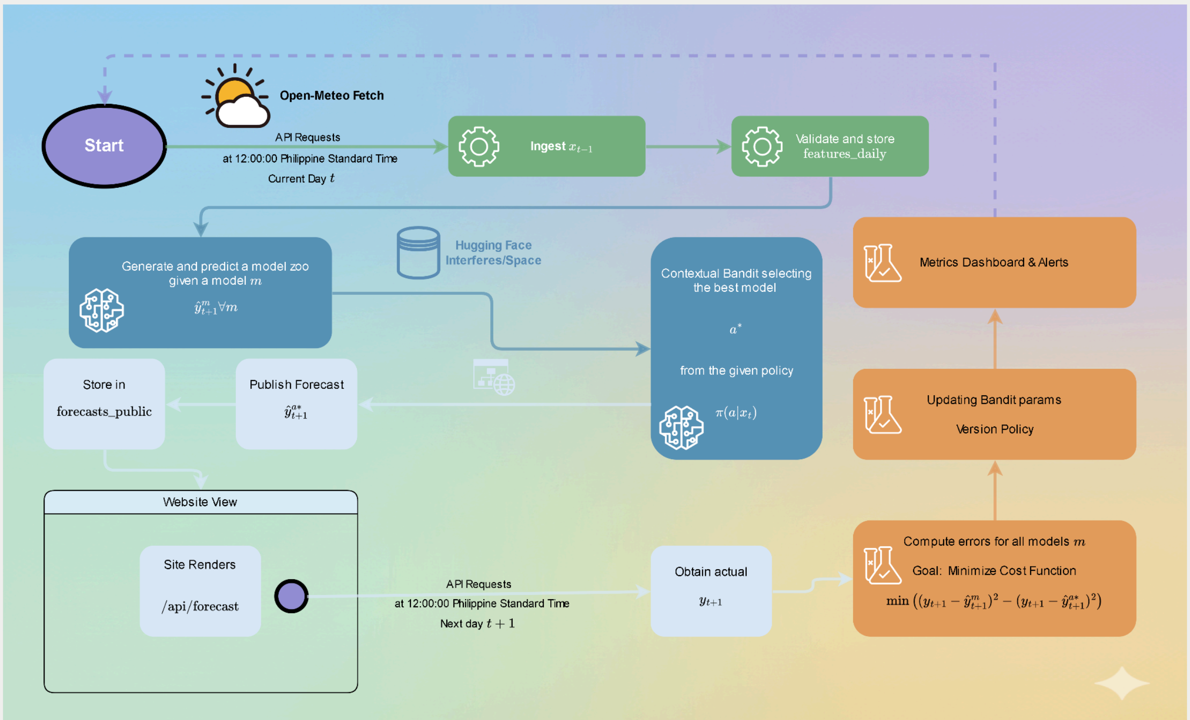
*Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.*

*Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.*

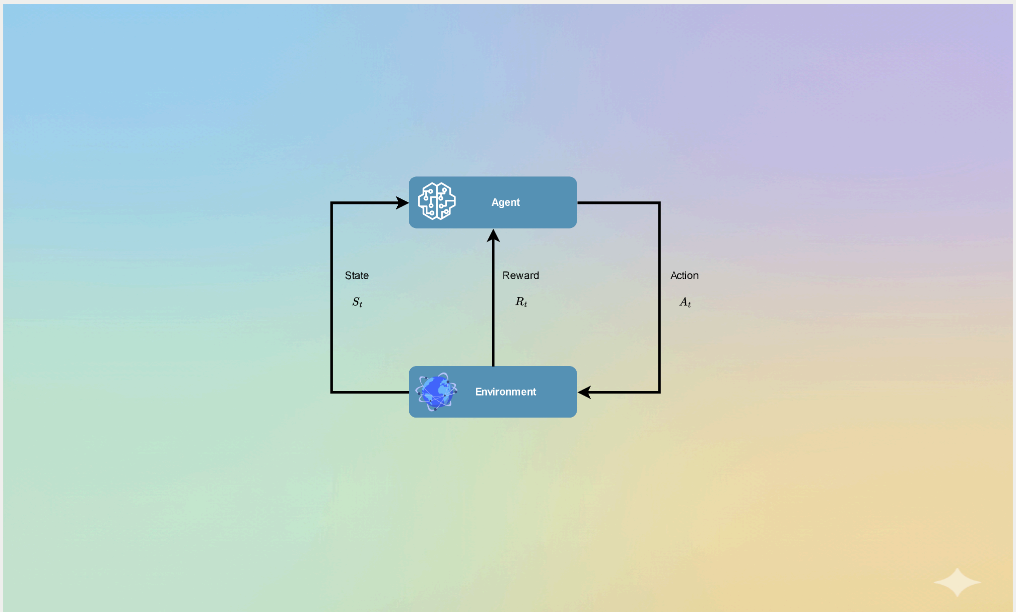
# APPENDICES




***Every day at 12:00*** *the orchestrator wakes up. It pulls today's weather signals from Open-Meteo and stamps them into Supabase. A lightweight model-zoo service on Hugging Face turns those signals into* ***tomorrow's HI predictions*** *for every candidate model. Our* ***bandit policy****—trained from months of logged outcomes—scores the context and* ***chooses one model****. That single forecast is published to the website and cached.*




***Tomorrow*** *when the true HI is reported, a second job computes errors for every model,* ***measures regret****, and* ***updates the policy****. The dashboard updates leaderboards and a rolling "regret to oracle" plot. If anything fails, a guardrail publishes a safe baseline forecast so the site never goes dark seamlessly.*




*This research utilizes a* ***model-free adaptive learning*** *methodology grounded in* ***contextual bandits*** *for* ***Adaptive Model Selection (AMS)****.*