AlphaSearch: towards Agentic AI for Price Arbitrage in Energy Systems

Millend Roy and Vladimir Pyltsov

Columbia University

INFORMS Annual Meeting · 2026

Based on our KDD 2026 paper

DOI: 10.1145/3770855.3818957

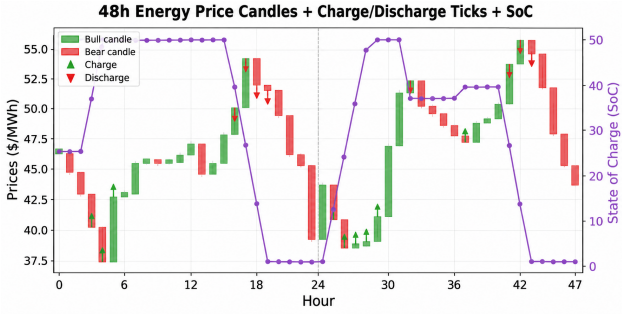
COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

IEOR Industrial Engineering & Operations Research COLUMBIA ENGINEERING

Quadracci SUSTAINABLE ENGINEERING LAB.

Motivation

Temporal arbitrage: Energy storage creates trading opportunities by shifting electricity usage across time.

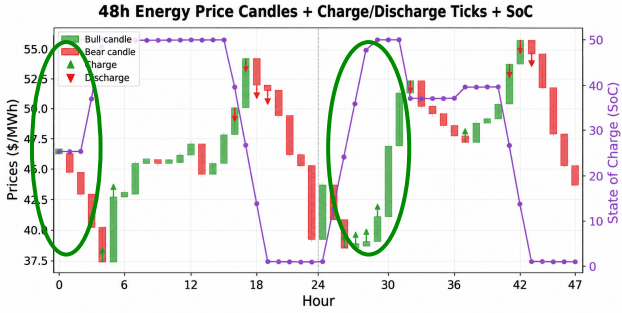


48-hour electricity prices and battery operation

Motivation

Temporal arbitrage: Energy storage creates trading opportunities by shifting electricity usage across time.

- **Charging:** Store electricity when prices are relatively low.

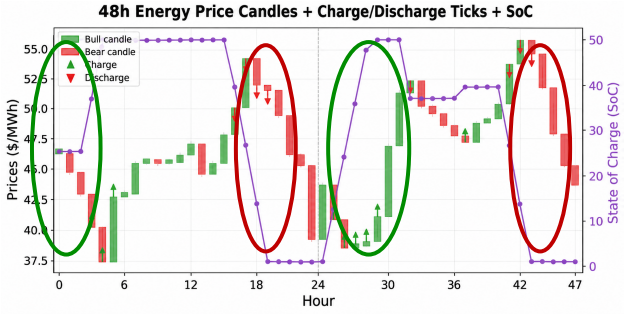


48-hour electricity prices and battery operation

Motivation

Temporal arbitrage: Energy storage creates trading opportunities by shifting electricity usage across time.

- **Charging:** Store electricity when prices are relatively low.
- **Discharging:** Sell stored energy when prices are relatively high.

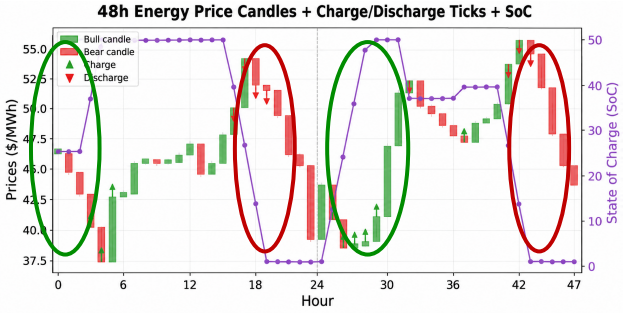


48-hour electricity prices and battery operation

Motivation

Temporal arbitrage: Energy storage creates trading opportunities by shifting electricity usage across time.

- **Charging:** Store electricity when prices are relatively low.
- **Discharging:** Sell stored energy when prices are relatively high.



48-hour electricity prices and battery operation

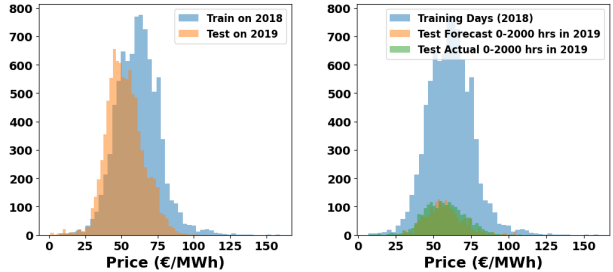
Note that Price volatility creates value but "***Why is this challenging***"?

Motivation: Why is this challenging?

- **Intertemporal coupling:** Each decision changes the energy available later.
- **Forecast uncertainty:** Future prices can differ from historical patterns.
- **Operational feasibility:** Charging, discharging and state of charge must obey physical constraints.

Motivation: Why is this challenging?

- **Intertemporal coupling:** Each decision changes the energy available later.
- **Forecast uncertainty:** Future prices can differ from historical patterns.
- **Operational feasibility:** Charging, discharging and state of charge must obey physical constraints.

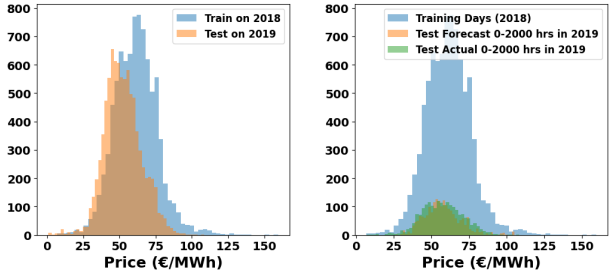


Historical and out-of-sample electricity price distributions in Italy

A schedule optimized for historical prices may perform poorly out-of-sample.

Motivation: Why is this challenging?

- **Intertemporal coupling:** Each decision changes the energy available later.
- **Forecast uncertainty:** Future prices can differ from historical patterns.
- **Operational feasibility:** Charging, discharging and state of charge must obey physical constraints.



Historical and out-of-sample electricity price distributions in Italy

A schedule optimized for historical prices may perform poorly out-of-sample.

Price Volatility creates value but ***Uncertainty Makes It Hard To Capture***.

Motivation: Can foundation models help?

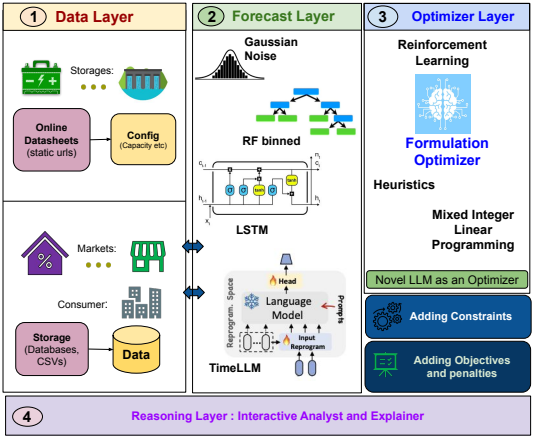
- Can pre-trained foundation models generalize to energy trading?
- Can **cross-modality adaptation** identify trading opportunities from time-series data using text prompts?
- Can reasoning capabilities provide interpretable explanations for scheduling decisions?

From price forecasting to decision-making

Forecasting $\longrightarrow$ Optimization $\longrightarrow$ Reasoning

How can we integrate these capabilities into a unified decision-support framework?

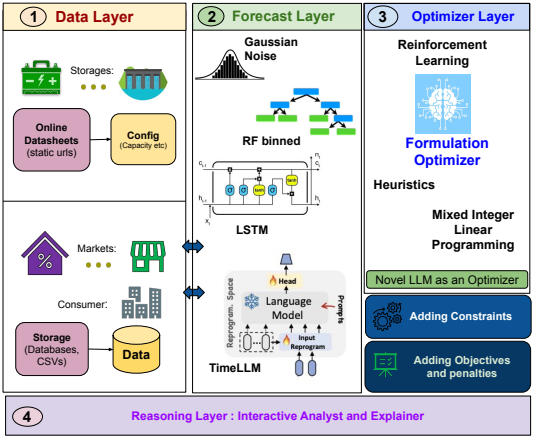
AlphaSearch: a modular, typed decision-support workflow



1. Modularity: Interchangeable models within the same pipeline:

- **Forecasting:** RF, LSTM, Time-LLM
- **Optimization:** MILP, PPO, LLMs

AlphaSearch: a modular, typed decision-support workflow



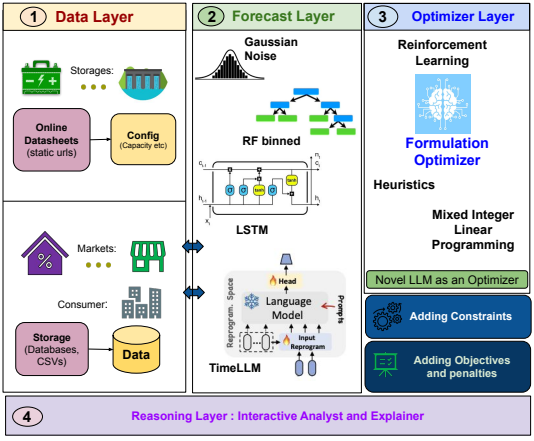
1. Modularity: Interchangeable models within the same pipeline:

- **Forecasting:** RF, LSTM, Time-LLM
- **Optimization:** MILP, PPO, LLMs

2. Typed Interfaces: Standardized data contracts between layers:

- Inputs: prices, forecasts, battery parameters
- Outputs: schedules, objective values, solver results

AlphaSearch: a modular, typed decision-support workflow



1. Modularity: Interchangeable models within the same pipeline:

- **Forecasting:** RF, LSTM, Time-LLM
- **Optimization:** MILP, PPO, LLMs

2. Typed Interfaces: Standardized data contracts between layers:

- Inputs: prices, forecasts, battery parameters
- Outputs: schedules, objective values, solver results

3. Feasibility Checks & Repair: Validate proposed schedules against:

- Charging/discharging power limits
- State-of-charge dynamics and bounds
- Initial and terminal constraints

Notations: Parameters, Inputs, and Decision Variables

Battery parameters θ^{bat}

C	battery capacity (MWh)
$P_c^{\max}, P_d^{\max}$	max charging / discharging power (MW)
η_c, η_d	charging / discharging efficiency
SoC_{init}	initial state of charge
$\text{SoC}_{\min}, \text{SoC}_{\max}$	operating SoC bounds
$\text{SoC}_{\text{target}}$	terminal SoC requirement

Daily market inputs ξ_t

π_t	realized electricity price at time t
$\hat{\pi}_t$	forecasted electricity price at time t
D_t	realized demand / load
$\hat{D}_t$	forecasted demand / load
X_t	additional exogenous market features

Decision variables

$P_c(t)$	Charging power at time t
$P_d(t)$	discharging power at time t
SoC_t	state of charge at time t

Problem Setup

Let $t = 0, \ldots, T-1$, with interval length Δt (hours).

1. Objective

$$\min_{\{P_c(t), P_d(t)\}_{t=0}^{T-1}} \sum_{t=0}^{T-1} \hat{\pi}_t P_{\text{net}}(t) \Delta t$$

2. Net grid purchase

$$P_{\text{net}}(t) = D_t + P_c(t) - P_d(t)$$

At each hour, the battery either increases grid purchases by charging, or reduces them by discharging.

For the hindsight oracle, replace forecast prices $\hat{\pi}_t$ with realized prices π_t.

3. Operational constraints

$$\begin{aligned} 0 \le P_c(t) &\le P_c^{\max} \\ 0 \le P_d(t) &\le P_d^{\max} \\ \text{SoC}_{\min} \le \text{SoC}_t &\le \text{SoC}_{\max} \\ \text{SoC}_0 &= \text{SoC}_{\text{init}} \\ \text{SoC}_T &\ge \text{SoC}_{\text{target}} \end{aligned}$$

4. State-of-charge dynamics

$$\text{SoC}_{t+1} = \text{SoC}_t + \frac{\eta_c P_c(t) \Delta t}{C} - \frac{P_d(t) \Delta t}{\eta_d C}$$

Benchmark 1: Deterministic Mixed-Integer Linear Programming

Additional MILP constraints

$$
\begin{aligned}
0 \le P_c(t) &\le P_c^{\max} y_t^c, \\
0 \le P_d(t) &\le P_d^{\max} y_t^d, \\
y_t^c + y_t^d &\le 1, \\
y_t^c, y_t^d &\in \{0,1\}.
\end{aligned}
$$

- $y_t^c = 1$: charging is permitted.
- $y_t^d = 1$: discharging is permitted.
- $y_t^c + y_t^d \le 1$ prevents simultaneous charging and discharging.

Two information settings

Forecast-based MILP
Optimize using $\hat{\pi}_t$.

Perfect-foresight MILP
Optimize using realized π_t.

The latter provides a hindsight benchmark, not a deployable trading policy.

But point forecasts may exhibit “systematic bias" and fail to capture uncertainty in future electricity prices.

Benchmark 2: Bias Correction & Stochastic MILP

1. Bias-corrected price forecasts: Estimate systematic forecast errors from historical data.

$$\widehat{b}_t = \frac{1}{N}\sum_{j=1}^{N}\left(\pi_t^{(j)} - \widehat{\pi}_t^{(j)}\right)$$

$$\widetilde{\pi}_t = \widehat{\pi}_t + \widehat{b}_t$$

2. Generate price scenarios: Perturb corrected forecasts using historical residuals.

$$\widehat{\pi}_t^{(s)} = \widetilde{\pi}_t + \varepsilon_t^{(s)}, \qquad s = 1, \ldots, S$$

How are historical residuals generated? For each historical day j, compute : $r_t^{(j)} = \pi_t^{(j)} - \widetilde{\pi}_t^{(j)}$

Sample an entire historical residual trajectory of a day:

$$J_s \sim \mathrm{Unif}\{1, \ldots, N\}$$
$$\varepsilon_t^{(s)} = r_t^{(J_s)}$$

3. Stochastic MILP objective: Select one day-ahead schedule across all scenarios.

$$\min_{x \in \mathcal{X}} \frac{1}{S}\sum_{s=1}^{S}\sum_{t=0}^{T-1}\widehat{\pi}_t^{(s)} P_{\mathrm{net}}(t; x)\Delta t$$

With a linear objective and one shared battery schedule, scenario-based expected-cost optimization depends only on average forecasted prices. **Explicit downside-risk modeling requires an additional risk-sensitive objective.**

Benchmark 3: Risk-Aware Optimization

1. Conditional Value-at-Risk (CVaR): Evaluate the same battery schedule under each price scenario.

$$L_s(x) = \sum_{t=0}^{T-1} \widehat{\pi}_t^{(s)} P_{\text{net}}(t; x) \Delta t$$

Measure the expected cost in the worst $1 - \beta$ fraction of scenarios. Introduce auxiliary variables z_s for the excess cost beyond threshold α.

$$z_s \geq L_s(x) - \alpha, \qquad z_s \geq 0$$

$$\text{CVaR}_\beta = \alpha + \frac{1}{(1-\beta)S} \sum_{s=1}^{S} z_s$$

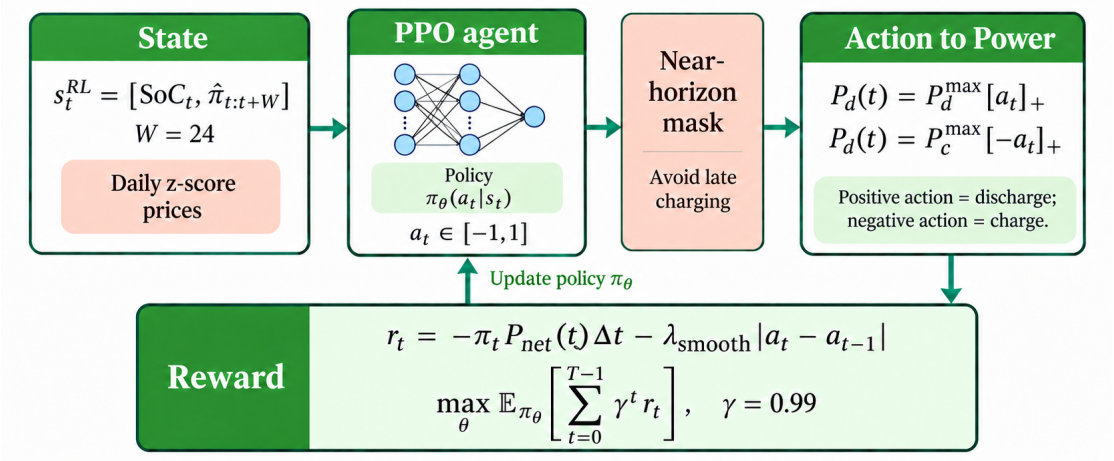
2. Risk-aware MILP objective: Balance expected operating cost and tail risk.

$$\min_{x \in \mathcal{X},\, \alpha,\, z} (1-\lambda) \frac{1}{S} \sum_{s=1}^{S} L_s(x) + \lambda \left[\alpha + \frac{\sum_{s=1}^{S} z_s}{(1-\beta)S} \right]$$

$\beta = 0.95$: worst 5% of scenarios; $\lambda = 0.5$: equal weighting of expected cost and CVaR.

Unlike the stochastic MILP, which optimizes **average scenario cost**, the CVaR formulation explicitly penalizes high-cost tail outcomes, yielding a more risk-sensitive battery schedule.

PPO: learn a sequential policy instead of solving a new MILP



PPO: A Physics-Informed No-Charge Action Mask

Define usable stored energy and remaining horizon:

$$E_t = C(\mathrm{SoC}_t - \mathrm{SoC}_{\min}), \qquad H = T - t$$

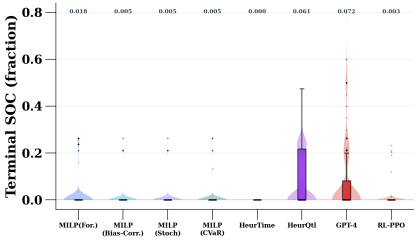
Sufficient condition to stop charging:

$$E_t \geq \frac{P_{\mathrm{d}}^{\max} \Delta t \, H}{\eta_{\mathrm{d}}}$$

Action masking: Positive charging actions are disabled when the sufficient condition is satisfied. $a_t \leftarrow \min(a_t, 0)$

Assumptions: Nonnegative remaining electricity prices and a nonbinding terminal SoC requirement.

End-of-Horizon Behavior



Terminal state-of-charge distribution across optimizers, Italy 2019.

LLM Scheduling: Prompting vs. Solver Distillation

1. Qwen: Supervised solver distillation

Training pairs
$x_i = (\hat{\pi}_i, \hat{D}_i, \theta^{\text{bat}})$, $y_i^* = \text{SoC}_{0:T,i}^*$

↓

Oracle supervision
y_i^* generated using MILP-oracle

↓

QLoRA fine-tuning
Frozen 4-bit base + trainable LoRA adapters

2. Inference: GPT-4 vs. Qwen

New instance
Forecasts + battery parameters

↓

GPT-4 zero-shot / Fine-tuned Qwen

↓

Candidate schedule $\widehat{\text{SoC}}_{0:T}$

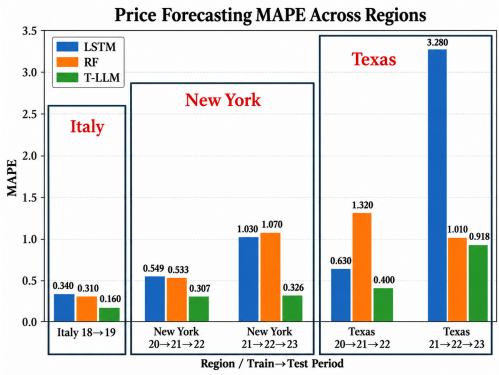
↓

Validation & repair
Physically feasible schedule

Supervised objective: $\mathcal{L}(\theta) = -\sum_i \sum_{j \in \mathcal{A}_i} \log p_\theta(y_{i,j} \mid x_i, y_{i,<j})$

Key distinction: Realized future prices generate offline oracle labels, but are not available to either LLM at inference.

Experiments: Evaluate Across Multiple Markets



- **Italy (Terna):** Relatively stable benchmark with lower forecasting errors.
- **New York (NYISO):** Different price regimes across rolling evaluation years.
- **Texas (ERCOT):** More challenging forecasting conditions, especially in 2023.

Forecasting comparison

Time-LLM achieves the **lowest MAPE across all five market-period comparisons**.

Does better forecasting accuracy translate into higher realized arbitrage profits?

Experiments (Italy): Single-Day vs. Long-Horizon Performance

4-HOUR BATTERY • $C = 43.13$ **MWh** • $P^{\max} = 10.78$ **MW** • $\eta_c = \eta_d = 0.95$ • $\mathrm{SoC}_0 = 0.5$

1. Single-Day Profit Comparison

Italy 2019: realized profit (€) across price forecasting methods

Method	Perfect	RF	LSTM	T-LLM
MILP	2246.09	1954.99	1901.09	2050.31
MILP (Bias-Corr.)	2246.17	1955.13	1983.45	2110.62
MILP (Stochastic)	2246.17	1958.88	1942.35	2111.43
MILP (CVaR)	2246.17	1980.15	1975.23	2080.71
Qwen (fine-tuned)	2245.76	**2191.13**	**2031.47**	**2130.14**
GPT-4	2093.80	1547.55	1417.72	1703.67
PPO	2009.78	1957.28	1952.05	2098.59

Qwen achieves **97.55% of oracle profit** and outperforms MILP by **12.1%** using RF forecasts on this representative day.

Experiments (Italy): Single-Day vs. Long-Horizon Performance

4-HOUR BATTERY • $C = 43.13$ **MWh** • $P^{\max} = 10.78$ **MW** • $\eta_c = \eta_d = 0.95$ • $\mathrm{SoC}_0 = 0.5$

1. Single-Day Profit Comparison

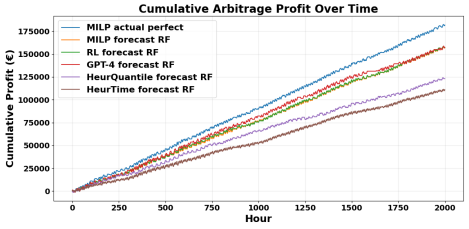
Italy 2019: realized profit (€) across price forecasting methods

Method	Perfect	RF	LSTM	T-LLM
MILP	2246.09	1954.99	1901.09	2050.31
MILP (Bias-Corr.)	2246.17	1955.13	1983.45	2110.62
MILP (Stochastic)	2246.17	1958.88	1942.35	2111.43
MILP (CVaR)	2246.17	1980.15	1975.23	2080.71
Qwen (fine-tuned)	2245.76	**2191.13**	**2031.47**	**2130.14**
GPT-4	2093.80	1547.55	1417.72	1703.67
PPO	2009.78	1957.28	1952.05	2098.59

Qwen achieves **97.55% of oracle profit** and outperforms MILP by **12.1%** using RF forecasts on this representative day.

2. Longer-Horizon Comparison

Italy 2019: cumulative realized profit over the first 2,000 hours



Forecast MILP, PPO, and GPT-4 maintain competitive cumulative profits, while the heuristics generally lag.

Experiments (Italy): Single-Day vs. Long-Horizon Performance

4-HOUR BATTERY • $C = 43.13$ **MWh** • $P^{\max} = 10.78$ **MW** • $\eta_c = \eta_d = 0.95$ • $\mathrm{SoC}_0 = 0.5$

1. Single-Day Profit Comparison

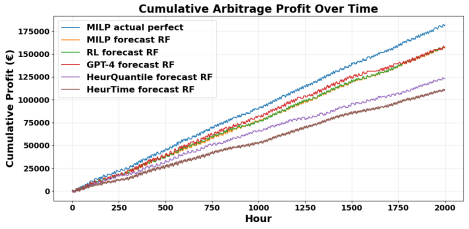
Italy 2019: realized profit (€) across price forecasting methods

Method	Perfect	RF	LSTM	T-LLM
MILP	2246.09	1954.99	1901.09	2050.31
MILP (Bias-Corr.)	2246.17	1955.13	1983.45	2110.62
MILP (Stochastic)	2246.17	1958.88	1942.35	2111.43
MILP (CVaR)	2246.17	1980.15	1975.23	2080.71
Qwen (fine-tuned)	2245.76	**2191.13**	**2031.47**	**2130.14**
GPT-4	2093.80	1547.55	1417.72	1703.67
PPO	2009.78	1957.28	1952.05	2098.59

Qwen achieves **97.55% of oracle profit** and outperforms MILP by **12.1%** using RF forecasts on this representative day.

2. Longer-Horizon Comparison

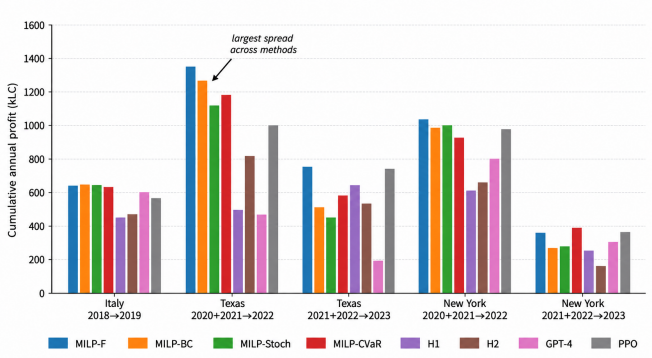
Italy 2019: cumulative realized profit over the first 2,000 hours



Forecast MILP, PPO, and GPT-4 maintain competitive cumulative profits, while the heuristics generally lag.

Strong single-day performance does not automatically imply long-horizon generalization.

Experiments: Does Forecast Accuracy Translate into Profit?



Cumulative realized profits across electricity markets (Italy: EUR; NY/TX: USD).

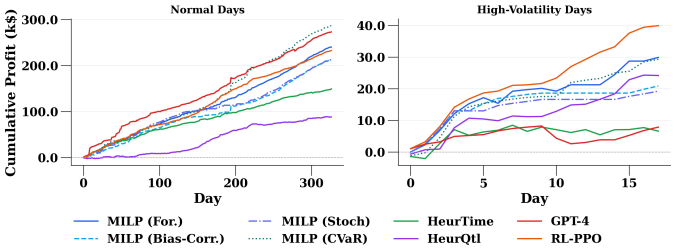
Key Findings

1. Forecast accuracy is not sufficient.
Lower price-forecasting errors do not guarantee higher arbitrage profits.

2. Performance varies by market.
Method rankings change across regions and evaluation years, highlighting sensitivity to market conditions.

The goal is not simply accurate forecasts, but profitable and robust decisions under uncertainty.

Experiments (NYISO): Performance Under Market Stress



Cumulative realized arbitrage profits under normal and high-volatility regimes.

1. Stress-regime identification

- Top 5% of days by maximum price.
- Top 10% of days by intraday price volatility.

2. PPO during market stress: achieves the **highest cumulative profit on high-volatility days** in New York 2023.

Strong overall profitability does not guarantee the best performance during market stress.

Takeaways and next steps

What we find

1. **Optimization remains essential:** MILP enforces exact operating constraints and yields strong long-horizon cumulative value.
2. **Learning can add value:** PPO offers competitive risk-adjusted performance; fine-tuned Qwen approaches the oracle in selected short-horizon tests.
3. **Forecasts are not enough:** evaluate end-to-end *realized dispatch profit*, not prediction accuracy alone.

Future Research agenda: degradation-aware, robust and multi-stage optimization with principled agentic orchestration and larger-scale evaluation.

Takeaways and next steps

What we find

1. **Optimization remains essential:** MILP enforces exact operating constraints and yields strong long-horizon cumulative value.
2. **Learning can add value:** PPO offers competitive risk-adjusted performance; fine-tuned Qwen approaches the oracle in selected short-horizon tests.
3. **Forecasts are not enough:** evaluate end-to-end *realized dispatch profit*, not prediction accuracy alone.

Future Research agenda: degradation-aware, robust and multi-stage optimization with principled agentic orchestration and larger-scale evaluation.

Thank you!

Scan to check our paper: