Fine-Tuning LLMs to Generate Economical and Reliable Actions for the Power Grid

Mohamad Chehade and Hao Zhu
Chandra Family Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX, USA
{chehade, haozhu}@utexas.edu

***Abstract*—Public Safety Power Shutoffs (PSPS) force rapid topology changes that can render standard operating points infeasible, requiring operators to quickly identify corrective transmission switching actions that reduce load shedding while maintaining acceptable voltage behavior. We present a verifiable, multi-stage adaptation pipeline that fine-tunes an instruction-tuned large language model (LLM) to generate *open-only* corrective switching plans from compact PSPS scenario summaries under an explicit switching budget. First, supervised fine-tuning distills a DC-OPF MILP oracle into a constrained action grammar that enables reliable parsing and feasibility checks. Second, direct preference optimization refines the policy using AC-evaluated preference pairs ranked by a voltage-penalty metric, injecting voltage-awareness beyond DC imitation. Finally, best-of-N selection provides an inference-time addition by choosing the best feasible candidate under the target metric. On IEEE 118-bus PSPS scenarios, fine-tuning substantially improves DC objective values versus zero-shot generation, reduces AC power-flow failure from 50% to single digits, and improves voltage-penalty outcomes on the common-success set. Code and data-generation scripts are released to support reproducibility.**

***Index Terms*—Public Safety Power Shutoff (PSPS), transmission switching, large language models, supervised fine-tuning, direct preference optimization, AC feasibility, voltage regulation**

I. Introduction

Large language models (LLMs) have rapidly transitioned from research prototypes to deployable decision-support tools across diverse domains [1]–[3]. Their ability to transform unstructured descriptions into structured outputs makes them attractive for operational environments where decisions are time-sensitive and consequences are high. Power system control rooms are a compelling setting: operators manage complex contingencies, coordinate actions across many assets, and balance reliability, economics, and compliance under tight time constraints [4]. Unlike traditional decision-support tools that require specialized inputs or rigid interfaces, LLMs enable operators to interact through natural language while producing machine-readable recommendations (e.g., structured action lists) that can be verified before execution [5], [6].

However, foundation LLMs lack domain-specific knowledge of power system physics, operational constraints, and grid safety requirements. Training grid-specific LLMs *from scratch* is impractical: modern LLMs succeed through pre-training on trillions of tokens spanning diverse domains [7], while grid operations data are orders of magnitude smaller and specialized. A practical alternative is to *adapt* a strong instruction-tuned model via targeted fine-tuning so that it can (i) read a compact, structured description of a grid scenario and (ii) output actions in a constrained grammar that can be checked for feasibility.

This work was funded by NSF Grant 2130706 and ARO Grant W911NF2310266.

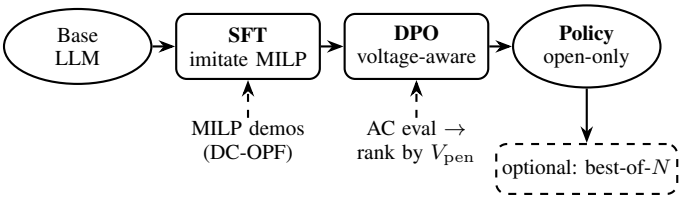


Fig. 1: Multi-stage adaptation pipeline for PSPS corrective switching.

In this work, we study a concrete and operationally motivated task: corrective, *open-only* transmission switching during Public Safety Power Shutoffs (PSPS), which are corrective de-energization actions used by utilities to reduce wildfire ignition risk during extreme weather conditions [8]. When PSPS forces lines out of service, operators must rapidly determine whether opening additional elements can mitigate overloads, reduce load shedding, and improve operating conditions while respecting switching budgets and operational rules. Computing optimal actions with mixed-integer optimization can be expensive under time pressure, especially when considering nonlinear AC constraints [9]–[11]. Our goal is to amortize this optimization effort into training, then produce high-quality switching recommendations at inference time using structured scenario summaries and a verifiable action grammar.

Figure 1 illustrates the pipeline we adopt. Starting from an instruction-tuned base model, supervised fine-tuning (SFT) trains the LLM to imitate MILP-derived open-only switching decisions under DC constraints. We then apply direct preference optimization (DPO) using ranked responses derived from AC voltage-quality evaluation, producing a voltage-aware policy that more reliably prioritizes actions with fewer voltage violations.

This design follows a standard alignment pattern for instruction-tuned LLMs: imitation learning first, followed by preference-based refinement [12], [13]. In our setting, the supervised stage anchors the policy to an optimization oracle,

while the preference stage injects AC voltage-awareness that is difficult to encode directly in DC training. The resulting model functions as a candidate-plan generator whose outputs can be parsed, verified, and evaluated with existing grid-analysis tools. Our contributions are:

- We formulate PSPS-aware open-only switching with switching budgets and corridor structure using a DC-OPF MILP oracle (Section II).
- We design a structured scenario representation and action grammar that enables an instruction-tuned LLM to emit switching plans that are straightforward to parse and verify (Section III).
- We introduce a voltage-aware preference refinement stage based on DPO, using AC-derived voltage-quality preferences to align the model beyond DC imitation (Section III-A).
- We evaluate economic performance, AC feasibility, and voltage quality, including comparisons to a neural baseline and training-curve reporting for reproducibility (Section IV).

Finally, we discuss practical considerations such as feasibility checks, training/inference costs, and deployment constraints (Section IV). We view this as a step toward verifiable, operator-facing LLM assistants that interface with existing grid analysis pipelines rather than replacing them.

II. PSPS Switching Problem

Public Safety Power Shutoffs (PSPS) are preventive and corrective de-energization actions taken by utilities to reduce wildfire ignition risk during extreme weather conditions [14], [15]. When a PSPS event forces a subset of transmission lines out of service, system operators must determine whether additional *corrective open-only* switching actions can improve reliability and reduce load shedding. We formulate a DC optimal power flow (DC-OPF) model that explicitly incorporates PSPS constraints and pose an open-only decision-making problem in which operators may proactively open a limited number of additional transmission elements to mitigate system stress.

a) Network model and PSPS constraints.: Consider a power system with bus set $\mathcal{B}$ ($|\mathcal{B}| = n_b$), transmission line set $\mathcal{E}$ ($|\mathcal{E}| = n_\ell$), and generator set $\mathcal{G} \subseteq \mathcal{B}$. Each line e has reactance $x_e > 0$ and rating $S_e^{\max} > 0$ (MW). Buses have demand $P_i^{\mathrm{d}} \geq 0$, generators have capacity $P_g^{\min} \leq P_g \leq P_g^{\max}$ and cost $c_g \geq 0$ (\$/MW), and load shedding $P_i^{\mathrm{s}} \geq 0$ is penalized at rate γ (\$/MW). A PSPS event is encoded by availability mask $\boldsymbol{\xi} \in \{0,1\}^{n_\ell}$, where $\xi_e = 0$ forces line e open. Operator decisions $z_e \in \{0,1\}$ determine effective status $g_e = \xi_e z_e$. We use standard DC power flow with bus-branch incidence A and susceptance $B_\ell = \mathrm{diag}(1/x_e)$.

b) Open-only switching with a line budget.: Operators may open up to K_ℓ additional PSPS-available lines. The open-only switching problem is:

$$\min_{\substack{\boldsymbol{P}_g, \boldsymbol{P}^{\mathrm{s}}, \boldsymbol{\theta}, \boldsymbol{P}_\ell, \\ \boldsymbol{z} \in \{0,1\}^{n_\ell}}} \quad \sum_{g \in \mathcal{G}} c_g P_g \; + \; \gamma \sum_{i \in \mathcal{B}} P_i^{\mathrm{s}} \tag{1a}$$

$$\text{s.t.} \quad P_g^{\min} \leq P_g \leq P_g^{\max}, \; \forall g \in \mathcal{G}, \tag{1b}$$

$$P_i^{\mathrm{s}} \geq 0, \; \forall i \in \mathcal{B}, \tag{1c}$$

$$\boldsymbol{P}_\ell = B_\ell A^\top \boldsymbol{\theta}, \; \theta_r = 0, \tag{1d}$$

$$A\boldsymbol{P}_\ell = \boldsymbol{P}_g - (\boldsymbol{P}^{\mathrm{d}} - \boldsymbol{P}^{\mathrm{s}}), \tag{1e}$$

$$-S_e^{\max} \xi_e z_e \leq P_{\ell,e} \leq S_e^{\max} \xi_e z_e, \; \forall e \in \mathcal{E}, \tag{1f}$$

$$z_e \in \{0,1\}, \; z_e = 0 \text{ if } \xi_e = 0, \; \forall e \in \mathcal{E}, \tag{1g}$$

$$\sum_{e \in \mathcal{E}} (1 - z_e)\, \xi_e \leq K_\ell. \tag{1h}$$

Constraint (1g) enforces that PSPS-forced outages remain open; (1h) limits operator-induced opens to K_ℓ available lines. This is structurally related to optimal transmission switching [9], [10] but restricted to open-only actions.

c) Corridor-constrained open-only switching.: Switching decisions may be constrained to transmission corridors $\mathcal{S}$—geographically grouped lines [15], [16]. Binary variables $y_S \in \{0,1\}$ indicate whether corridor S is activated for switching. Corridor and line decisions are coupled by:

$$z_e \geq 1 - y_S, \quad \forall S \in \mathcal{S}, \; \forall e \in S, \tag{2a}$$

$$\sum_{S \in \mathcal{S}} y_S \leq K_S, \quad y_S \in \{0,1\}, \; \forall S \in \mathcal{S}, \tag{2b}$$

When $y_S = 0$, (2a) prevents operator opens in corridor S; when $y_S = 1$, line decisions remain free. Constraint (2b) limits activated corridors to K_S.

d) Role in the pipeline.: We use the DC-OPF MILP above as an *offline oracle* to generate labeled switching plans for supervised adaptation. Voltage quality and AC feasibility are evaluated separately using AC power flow in the experimental section, enabling us to train on DC structure while assessing voltage-critical performance.

III. Supervised Fine-Tuning (SFT) LLMs

Solving the mixed-integer linear program (MILP) in (1) for every PSPS scenario can be computationally expensive, particularly when evaluating large numbers of contingencies or when operators require rapid what-if analysis. Similar scalability challenges are well documented for optimal transmission switching (OTS) formulations, motivating learning-assisted and proxy-based approaches that approximate an optimizer while preserving most of the economic benefit [17], [18]. We therefore train a large language model (LLM) to *imitate* an optimization oracle, amortizing the cost of offline MILP solves into a single supervised fine-tuning (SFT) phase. After SFT, the model generates candidate switching plans from compact scenario summaries, which are then parsed and verified before evaluation or deployment. This section describes the oracle, the input–output representation, and the SFT protocol.

a) Ground-Truth Oracle.: Given PSPS mask $\boldsymbol{\xi}$ and budget K_ℓ, we solve the DC-OPF MILP (1) to obtain optimal operator-induced opens

$$\mathcal{T}(\boldsymbol{\xi}) \; \triangleq \; \big\{\, e \in \mathcal{E} : \xi_e = 1, \; z_e^\star = 0 \,\big\}. \tag{3}$$

For corridor-constrained problems, we solve the variant with (2).

b) Scenario Representation.: Each training example corresponds to a PSPS scenario with mask $\boldsymbol{\xi}$ and budgets K_ℓ, K_S. We provide a structured textual summary including case dimensions, the number of PSPS-forced opens, and a corridor breakdown listing member lines with their forced/eligible status and applicable budgets. This abstracts away numerical parameters while preserving topological structure.

c) Action Grammar.: The target encodes $\mathcal{T}(\boldsymbol{\xi})$ using: `open(Sk:LINE)` for corridor-associated lines (e.g., `open(S6:135)`), `open(LINE)` for non-corridor lines, and `do_nothing` if $\mathcal{T}(\boldsymbol{\xi}) = \emptyset$. We sort actions lexicographically to form a canonical string, e.g.,

```
open(S3:21); open(S7:98); open(131).
```

d) Dataset Formatting.: Supervised pairs use a chat format: system prompt defining the task and grammar, user message with scenario JSON, and assistant message with the canonical action string, following standard instruction-tuning practice [12].

e) Fine-Tuning Objective and Protocol.: We initialize from an instruction-tuned base LLM and perform supervised fine-tuning on the training partition to learn a policy π_{SFT} that maps scenario summaries to action strings. The objective is the standard conditional language-modeling loss over the assistant tokens. Let s_k denote the scenario summary and a_k the canonical action string for example k. We optimize parameters ϕ by minimizing

$$\mathcal{L}_{\mathrm{SFT}}(\phi) \;=\; -\sum_{k} \log p_\phi\big(a_k \,\big|\, s_k,\ \text{system prompt}\big), \tag{4}$$

where the sum ranges over training examples. This mirrors the supervised stage used in instruction-following pipelines prior to preference optimization [12]. The resulting model π_{SFT} is then used for inference and (optionally) subsequent preference-based refinement.

f) Parsing and Verification.: At inference, we parse outputs and verify: (i) grammar validity, (ii) all opened lines are PSPS-available ($\xi_e = 1$), and (iii) budget adherence. Invalid outputs are flagged, ensuring the LLM functions as a candidate generator rather than an unverified controller.

A. Improving Voltage Safety via DPO (SFT-Time)

The supervised policy π_{SFT} imitates a DC open-only oracle, but DC imitation alone does not explicitly optimize voltage quality under nonlinear AC physics. To better align the policy with voltage safety objectives, we introduce a refinement stage based on *direct preference optimization* (DPO) [13]. DPO is an reinforcement learning-free objective that trains a policy from pairwise preferences (y^+, y^-), encouraging the refined policy to assign higher probability to actions that yield better voltage outcomes under AC evaluation.

a) Voltage-Penalty Metric.: For candidate plan y, we parse into topology $\boldsymbol{z}(y)$ and solve AC power flow to obtain bus voltage magnitudes $\{|V_i(y)|\}$. We define a deadband v_{db} around nominal voltage and penalize violations outside $[1 - v_{\mathrm{db}},\, 1 + v_{\mathrm{db}}]$:

$$V_{\mathrm{pen}}(y) \;\triangleq\; \kappa \sum_{i \in \mathcal{B}} \Big(\max\big\{\, \big||V_i(y)| - 1\big| - v_{\mathrm{db}},\ 0 \,\big\} \Big)^p, \tag{5}$$

where $p \in \{1, 2\}$ selects aggregation norm and $\kappa > 0$ scales the penalty. Non-convergent AC solutions receive penalty V_{fail} (large constant).

b) Preference Pair Construction.: For each scenario x, we sample N_{pref} candidates from $\pi_{\mathrm{SFT}}(\cdot \mid x)$, discard malformed or budget-violating plans, evaluate $V_{\mathrm{pen}}(y)$ via AC power flow, and select (y^+, y^-) pairs where

$$V_{\mathrm{pen}}(y^-) - V_{\mathrm{pen}}(y^+) \geq \Delta_{\mathrm{pref}}, \tag{6}$$

yielding preference dataset $\mathcal{D}_{\mathrm{volt}} = \{(x_i, y_i^+, y_i^-)\}_{i=1}^M$ where y^+ is preferred for voltage quality.

c) DPO Objective.: Let π_ϕ denote the trainable policy initialized from π_{SFT}, with reference $\pi_{\mathrm{ref}} = \pi_{\mathrm{SFT}}$. For triple (x, y^+, y^-), define

$$\Delta_\phi(x, y^+, y^-) \;\triangleq\; \log \pi_\phi(y^+ \mid x) - \log \pi_\phi(y^- \mid x), \tag{7}$$

and analogously Δ_{ref}. The DPO loss is

$$\mathcal{L}_{\mathrm{DPO}}(\phi) \;=\; -\sum_{(x, y^+, y^-) \in \mathcal{D}_{\mathrm{volt}}} \log \sigma\Big(\beta_{\mathrm{DPO}} \big[\Delta_\phi - \Delta_{\mathrm{ref}} \big] \Big), \tag{8}$$

where $\beta_{\mathrm{DPO}} > 0$ controls preference strength [13]. Minimizing this produces π_{DPO}, which increases likelihood of low-penalty plans while anchored to the SFT reference.

d) Scope and Practicality.: Preference construction uses AC power flow *offline* to label candidates with voltage-quality information. At inference time, the refined model can be used either in single-shot mode or combined with best-of-N reranking (Section III-B), where V_{pen} can serve as a selection criterion when AC evaluation is available.

B. Improving Voltage Safety via Best-of-N (Inference-Time)

Even after supervised fine-tuning and preference alignment, a single stochastic decode can produce a suboptimal or malformed plan. We therefore use *best-of-N* inference as a lightweight, training-free mechanism to improve solution quality by sampling multiple candidate plans and selecting the best under a task metric. Best-of-N is widely used in reasoning and structured prediction tasks to exploit sampling diversity, often with strong gains at moderate N [19].

Given a scenario summary x and policy π (e.g., π_{SFT} or π_{DPO}), we draw N independent candidates:

$$y^{(j)} \sim \pi(\cdot \mid x), \qquad j = 1, \ldots, N. \tag{9}$$

Each candidate is verified through staged checks: grammar parsing, PSPS/budget constraint validation, DC feasibility evaluation, and optional AC power flow for voltage scoring. Let $\mathcal{Y}_{\mathrm{valid}}(x)$ denote candidates passing verification. If empty, we fall back to a safe default (e.g., doing nothing). We then select the plan minimizing a scalar score:

$$\hat{y}(x) \;\in\; \arg\min_{y \in \mathcal{Y}_{\mathrm{valid}}(x)} \mathrm{Score}(x, y). \tag{10}$$

Primary choices are: (i) DC economic score $J_{\mathrm{DC}}(x, y)$, or (ii) AC voltage penalty $V_{\mathrm{pen}}(y)$ from (5). When both matter, we use scalarization $\mathrm{Score}(x, y) = J_{\mathrm{DC}}(x, y) + \lambda V_{\mathrm{pen}}(y)$ with $\lambda \geq 0$ set by operational priorities.

Best-of-N increases compute linearly in N. We use moderate N with early stopping when target scores are met. Sampling is embarrassingly parallel, and N can be adjusted to trade off compute for quality. We quantify costs and inference times in Section IV.

IV. Experimental Results

Experiments use the IEEE 118-bus test system from MATPOWER [20] ($n_b = 118$, $n_\ell = 186$, $n_g = 54$). We define $|\mathcal{S}| = 9$ transmission corridors by grouping geographically proximate lines (8–20 lines per corridor). PSPS events are simulated by selecting corridor lines to force open, producing availability mask $\boldsymbol{\xi}$ for each scenario. Unless stated otherwise, operators may open at most $K_\ell = 3$ additional PSPS-available lines.

a) Datasets.: For SFT, we use 200 PSPS scenarios split between training and held-out testing. For DPO, we construct 440 preference pairs (x, y^+, y^-) by sampling candidates from π_{SFT} and ranking by AC voltage penalty V_{pen} (Section III-A). Data-generation scripts and splits are in the released codebase.

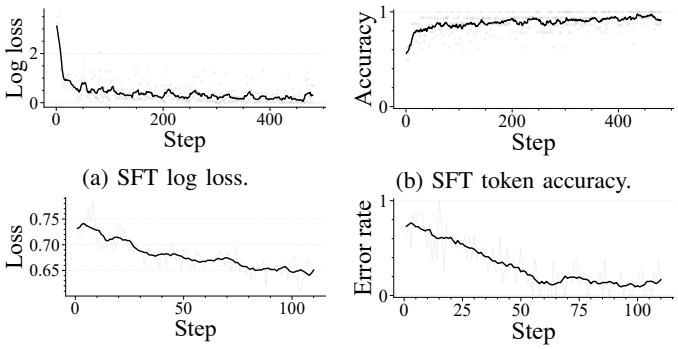
b) Implementation and Models.: Power flow and optimization use MATLAB R2024b with YALMIP [21]. We adapt an instruction-tuned LLM, `ft:gpt-4.1-mini-2025-04-14`, via the OpenAI fine-tuning API. SFT trains for 3 epochs (batch size 1, LR multiplier 2) on 453,384 training tokens. DPO initializes from π_{SFT} and trains for 2 epochs (batch size 8) with $\beta_{\mathrm{DPO}} = 0.1$ on 1,611,736 tokens. For the evaluation of the voltage penalties, we use $v_{\mathrm{db}} = 0$, $p = 1$, $\kappa = 1$ for (5). Our pipeline is backend-agnostic and can be applied to any instruction-tuned LLM deployed either via API or locally hosted models.

c) Compared Policies.: We compare: (i) zero-shot base LLM, (ii) π_{SFT}, (iii) π_{DPO}, and (iv) a neural network baseline: a fully-connected MLP with one hidden layer of width 512 (ReLU), trained on the same supervised dataset to predict corrective opens under identical budget and feasibility constraints.

d) Code Release.: Full code is available on GitHub: MFHChehade/LLM-Grid-Actions.

A. Training Curves for the Fine-Tuning Process

Figures 2(a–b) show that the SFT process is well-behaved: log loss drops sharply early then decreases gradually, while token accuracy rapidly increases and stabilizes. This pattern reflects the model learning the output grammar then refining scenario-to-action mappings. Figure 2(c–d) shows stable DPO optimization with decreasing loss and preference error rate, indicating successful separation of preferred vs. dispreferred plans. No divergence occurs under the chosen β_{DPO} and dataset size.



(a) SFT log loss. (b) SFT token accuracy.

(c) DPO loss. (d) DPO error rate.

Fig. 2: Training curves from the fine-tuning jobs (SFT and DPO) show the convergence.

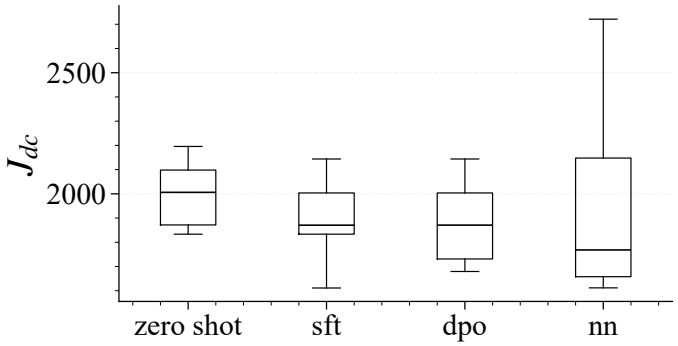


Fig. 3: Distribution of DC objective J_{DC} across all four compared policies (zero-shot, SFT, DPO, and NN).

B. DC Objective

Figure 3 reports the box plots for the J_{DC} distributions. Both SFT and DPO shift downward relative to zero-shot, distilling the oracle signal into lower-cost decisions. SFT and DPO medians are close, showing voltage-aware preference refinement preserves DC performance. The NN baseline achieves low median J_{DC} but exhibits wider dispersion and heavier upper tail, indicating occasional high-cost decisions.

C. AC Feasibility and Voltage Quality

Figure 4 shows zero-shot fails AC power flow in half of scenarios, while SFT and DPO reduce failures to a small fraction. This indicates DC oracle imitation plus constrained grammar substantially improves physical plausibility. The NN baseline achieves the lowest failure rate.

We evaluate V_{pen} on the *common-success set* (Figure 5)—scenarios where SFT, DPO, and NN all converge. DPO achieves lower median V_{pen} than SFT and NN, consistent with preference refinement improving voltage outcomes beyond DC imitation. However, the upper tail persists across policies, suggesting some scenarios remain challenging for voltage regulation and motivating richer preference data as future work.

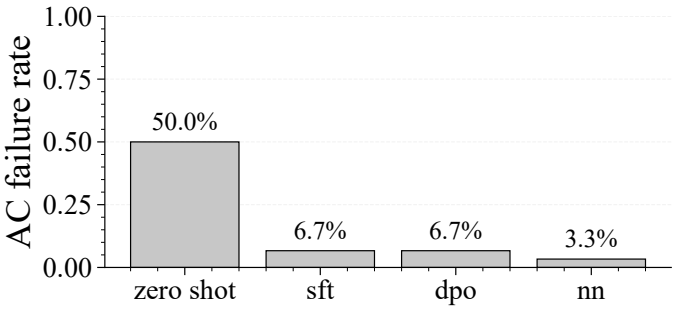


Fig. 4: AC power-flow failure rate across compared policies. Fine-tuned policies drastically reduce AC failures relative to the zero-shot baseline.

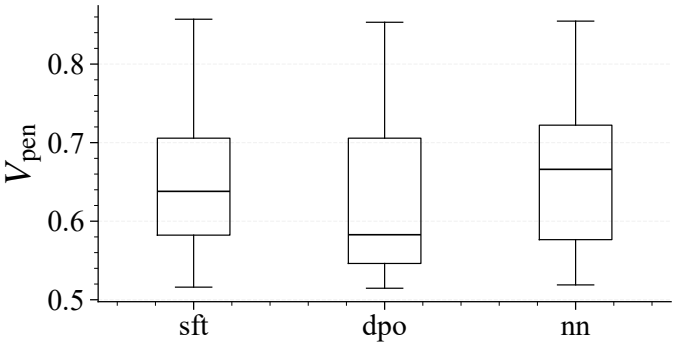


Fig. 5: Voltage penalty V_{pen} distribution on the *common-success set*, i.e., scenarios where all compared policies (SFT, DPO, NN) have achieved AC convergence. This ensures an apples-to-apples comparison of voltage quality.

V. Conclusion

We developed a verifiable fine-tuning pipeline that adapts a foundation LLM into a corrective switching assistant for PSPS events. Supervised fine-tuning distills MILP-derived DC-optimal *open-only* actions into a constrained, parseable grammar that enables systematic feasibility checks. Direct preference optimization then injects AC voltage-awareness by learning from preference pairs ranked using a voltage-penalty metric, improving voltage quality among AC-feasible cases. Best-of-N selection complements training by trading inference compute for improved candidate quality. Empirically on IEEE 118-bus PSPS scenarios, the fine-tuned policies outperform zero-shot generation in economic objective, dramatically reduce AC infeasibility, and yield improved voltage-penalty distributions on the common-success set, while remaining compatible with standard power-flow verification. Future work will investigate multi-task fine-tuning of a single foundation model so it can support diverse power-grid applications under a unified, verifiable decision framework.

References

[1] A. Mastropaolo, L. Pascarella, E. Guglielmi, M. Ciniselli, S. Scalabrino, R. Oliveto, and G. Bavota, "On the robustness of code generation techniques: An empirical study on GitHub Copilot," in *IEEE/ACM International Conference on Software Engineering (ICSE)*, 2023, also available as arXiv:2302.00438.

[2] Y. Zhang, S. A. Khan, A. Mahmud, H. Yang, A. Lavin, M. Levin, J. Frey, J. Dunnmon, J. Evans, A. Bundy, S. Džeroski, J. Tegnér, and H. Zenil, "Exploring the role of large language models in the scientific method: From hypothesis to discovery," *npj Artificial Intelligence*, vol. 1, p. 14, 2025.

[3] J. Lai, X. Li, Z. Wang, Y. Zhu, Y. Li, and S. Wang, "Large language models in law: A survey," *AI Open*, vol. 6, pp. 181–196, 2024.

[4] C.-J. Lin, M. M. Robertson, and J. D. Lee, "Analyzing the staffing and workload in the main control room," *Safety Science*, vol. 57, pp. 161–168, 2013.

[5] S. Kalami *et al.*, "Powering the grid with language: How LLMs are transforming energy systems," Medium (online article), 2025, accessed 2025.

[6] M. Sanguinetti, L. Atzori, R. Girau, and M. Marras, "Conversational agents for energy awareness and efficiency: A survey," *Electronics*, vol. 13, no. 2, p. 401, 2024.

[7] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," *arXiv preprint*, 2020.

[8] Pacific Gas and Electric Company, "Public safety power shutoffs (PSPS) fact sheet," Website, 2024. [Online]. Available: https://www.pge.com/

[9] E. B. Fisher, R. P. O'Neill, and M. C. Ferris, "Optimal transmission switching," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1346–1355, 2008.

[10] K. W. Hedman, R. P. O'Neill, E. B. Fisher, and S. S. Oren, "Optimal transmission switching with contingency analysis," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1577–1586, 2009.

[11] E. Haag, N. Rhodes, and L. Roald, "Long solution times or low solution quality: On trade-offs in choosing a power flow formulation for the optimal power shutoff problem," *Electric Power Systems Research*, vol. 234, p. 110713, 2024, also available as arXiv:2310.13843.

[12] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," in *Advances in Neural Information Processing Systems*, 2022, also known as the InstructGPT paper.

[13] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn, "Direct preference optimization: Your language model is secretly a reward model," in *International Conference on Learning Representations (ICLR)*, 2024.

[14] N. Rhodes, L. Ntaimo, and L. Roald, "Balancing wildfire risk and power outages through optimized power shut-offs," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3118–3128, 2021.

[15] A. Kody, R. Piansky, and D. K. Molzahn, "Optimizing transmission infrastructure investments to support line de-energization for mitigating wildfire ignition risk," in *Proceedings of the 11th Bulk Power Systems Dynamics and Control Symposium (IREP)*, 2022, also available as arXiv:2203.10176.

[16] A. Moreira, F. Pianco, B. Fanzeres, A. Street, R. Jiang, C. Zhao, and M. Heleno, "Distribution system operation amidst wildfire-prone climate conditions under decision-dependent line availability uncertainty," *IEEE Transactions on Power Systems*, 2024, also available as arXiv:2306.06336.

[17] S. Pineda, J. M. Morales, and A. Jiménez-Cordero, "Learning-assisted optimization for transmission switching," *TOP*, vol. 32, pp. 489–516, 2024.

[18] A.-A. B. Bugaje, J. L. Cremer, and G. Strbac, "Real-time transmission switching with neural networks," *IET Generation, Transmission & Distribution*, vol. 17, no. 3, pp. 696–705, 2023.

[19] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. H. Chi, S. Narang, A. Chowdhery, and D. Zhou, "Self-consistency improves chain-of-thought reasoning in language models," in *International Conference on Learning Representations (ICLR)*, 2023, also available as arXiv:2203.11171.

[20] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011. [Online]. Available: https://matpower.org/

[21] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proceedings of the IEEE International Symposium on Computer Aided Control System Design (CACSD)*, Taipei, Taiwan, 2004, pp. 284–289.