Power-Q: A Hybrid Screener-Analyst Framework Using LLM Agents for Real-Time Grid Monitoring and Diagnostics

Furqan Amjad, Tarmo Korõtko, Argo Rosin

Department of Electrical Power Engineering and Mechatronics, Tallinn University of Technology, Estonia

Motivation

- **Problem:** Modern microgrids with renewable generation are more dynamic, less inertial, and more exposed to power-quality disturbances.
- Existing monitoring and AI-based classification tools often detect the disturbance type, but they usually do not provide enough operational context, such as the affected bus, probable source, propagation path, or recommended operator response.
- **Core idea:** Power-Q combines fast deterministic screening agents with LLM-based analyst agents to transform raw voltage events into structured, explainable, topology-aware diagnostic information.
- **Why this work is needed:** Operators require decision-ready information, not only alarms or class labels. A human-in-the-loop diagnostic layer can reduce cognitive burden and improve situational awareness during fast-evolving grid events.
- The framework is intended for the supervisory monitoring layer, supporting operator interpretation and decision-making rather than replacing protection relays or control mechanisms.

Objective and questions

- **Q1 (Latency & Feasibility):**How does a hybrid MAS combining deterministic screening with LLM-based reasoning perform against the latency constraints?
- **Q2 (Standards Alignment):** To what degree do generative AI agents generate diagnostic reports that remain consistent with established power quality standards, while avoiding hallucination?
- **Q3 (Operator Utility):** To what extent can an LLM-based "Ticker" and "Query" system improve situational awareness?

Hybrid framework

1. **Screening agents** monitor wind, solar, and load domains using deterministic voltage thresholds.
2. **Orchestrator** aggregates event flags and dispatches structured context only when needed.
3. **LLM analyst agents** generate live summaries, diagnostic reports, and RAG-based operator answers.

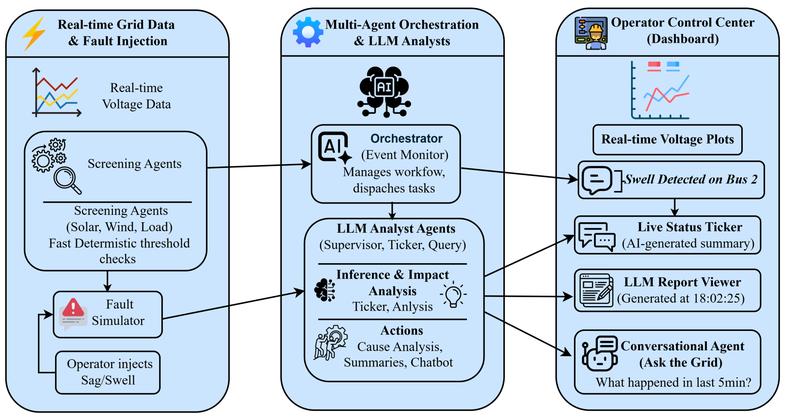

Figure 1. High-level Power-Q architecture showing the workflow from real-time monitoring and fault injection to multi-agent LLM-based diagnosis and operator support.

Agent functions

Table 1. Functional roles of the Power-Q agents.

Component	Operational function
Screeners	Fast sag/swell detection using domain-specific threshold checks.
Orchestrator	Event monitoring, context packaging, severity filtering, and task dispatch.
Ticker Agent	One-sentence, operator-readable event summaries.
Supervisor Agent	Root-cause interpretation, impact assessment, and recommendations.
Query Agent	RAG-grounded natural language access to recent event history and topology.

Acknowledgements

This work was supported by the project "Increasing the Knowledge Intensity of Ida-Viru Entrepreneurship", co-funded by the European Union under Grant 2021-2027.6.01.23-0034.

Detailed workflow

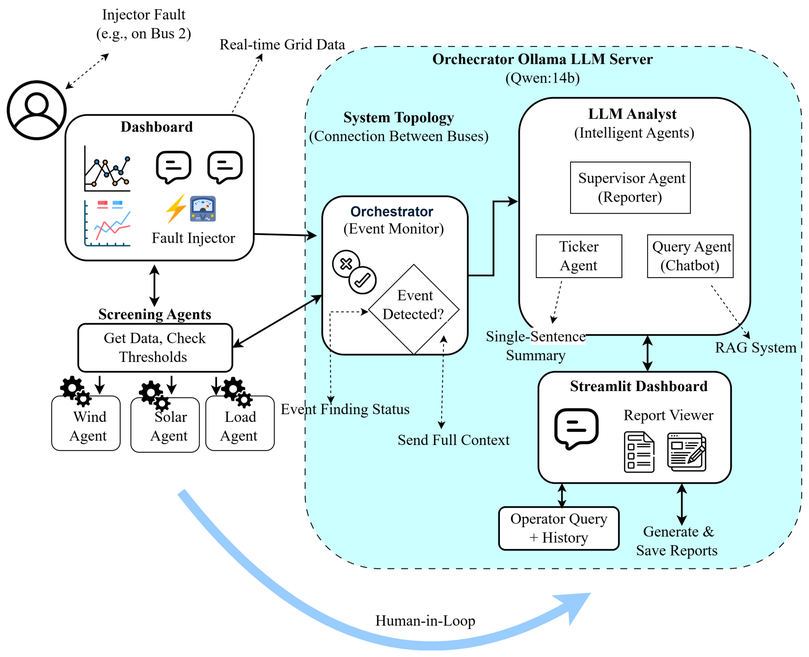

Figure 2. Detailed functional architecture of the Power-Q multi-agent system, showing the closed-loop data flow from decentralized screening agents to the LLM-based reasoning engine and human-in-the-loop operator interface.

Case study: simulated 5-bus microgrid

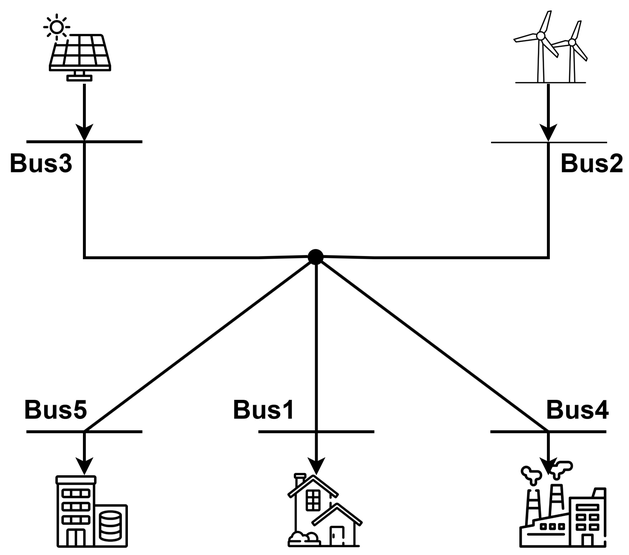

Figure 4. Test feeder microgrid.

Detection logic

$$\text{Sag}: \quad \frac{V_{\min}}{V_{\text{nom}}} < 0.9, \quad t \geq 0.5 \text{ cycle}$$

$$\text{Swell}: \quad \frac{V_{\max}}{V_{\text{nom}}} > 1.1, \quad t \geq 0.5 \text{ cycle}$$

Cascading wind-farm fault scenario

- **T0:** short-circuit-like shunt injected at Bus 2.
- **T0 to T0+1 s:** Bus 2 falls to about 0.85 p.u.; Bus 4 falls to about 0.89 p.u.; Bus 1 moves to about 0.92 p.u.; Buses 3 and 5 remain nominal.
- **T0+1.5 s:** Ticker Agent reports the sag and correlated Bus 4 impact.
- **T0+3.5 s:** Supervisor Agent attributes the primary cause to Bus 2 and flags Bus 4 as a secondary impacted load.

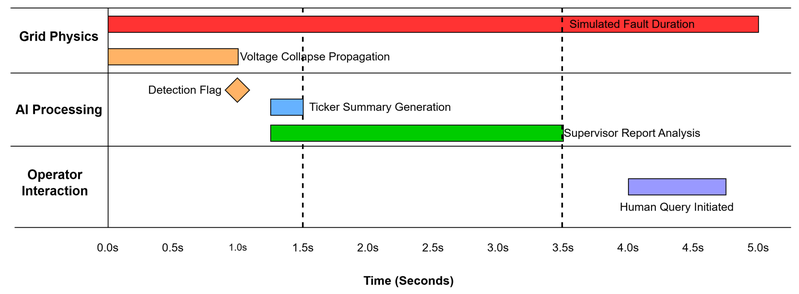

Figure 3. Cascading disturbance timeline.

Future work

- Validate Power-Q on larger feeders, noisy measurements, and a broader set of disturbance scenarios.
- Extend the evidence layer with current phasors, frequency, waveform features, and confidence scoring.
- Evaluate lightweight or domain-adapted LLM variants for practical edge or local deployment.
- Conduct operator-in-the-loop studies to assess usability, workload reduction, and decision quality.

Quantitative performance

Table 2. Latency performance of the Power-Q diagnostic pipeline.

Metric	Average	Std. dev.	Interpretation
Initial detection	185 ms	35 ms	deterministic screening
Ticker summary	1950 ms	210 ms	short LLM summary
Query response	2200 ms	280 ms	RAG operator answer
Supervisor report	3650 ms	450 ms	full diagnostic synthesis

Standards alignment and evaluation

Table 3: Evaluation outcomes reported in the Power-Q paper.

Evaluation aspect	Reported outcome
Supervisor standards alignment	Diagnostic reports aligned with key reporting elements from IEEE 1159 and IEEE C37.114, including event classification, timing, location, cause, impact, and mitigation information.
Root-cause interpretation	LLM-generated diagnostic interpretation was consistent with the expert-defined ground truth for fault location, source identification, and propagation path in the evaluated scenarios.
Ticker event coverage	The Ticker Agent summarized all ten injected power-quality events, including five voltage sags and five voltage swells, with no omission or false detection.
Ticker readability	Generated ticker summaries were clear and operator-readable, with only minor phrasing issues reported in a few cases.
Query contextual accuracy	The Query Agent achieved 100% factual accuracy in temporal, spatial, and negative-context retrieval tests under the evaluated controlled scenarios.

Pros of Power-Q

Table 4. Qualitative comparison between a rule-based expert system and Power-Q.

Capability	Rule-based base-line	Power-Q
Detection	Threshold rules	Threshold rules
Root-cause interpretation	Fixed heuristics	Contextual, topology-aware reasoning
Multi-bus events	Limited predefined mappings	Dynamic multi-source synthesis
Explanation	Static templates	Structured, adaptive reports
Operator interaction	Not supported	Natural language Query Agent
Scenario adaptation	Manual rule updates	Prompt/context-based reasoning

Main findings

- The hybrid screener-analyst architecture separated fast deterministic event detection from slower LLM-based interpretation, enabling rapid anomaly flagging while preserving explainable supervisory diagnostics.
- In the cascading wind-farm fault scenario, Power-Q identified the primary disturbance at Bus 2 and interpreted the voltage sag at Bus 4 as a correlated secondary impact.
- The deterministic screening layer achieved sub-200 ms initial detection, while LLM-based ticker summaries, operator queries, and supervisor reports remained within a second-level supervisory decision-support timescale.
- The Supervisor Agent generated structured diagnostic reports containing event classification, timing, location, cause identification, impact assessment, and mitigation-oriented information.
- The Ticker and Query agents improved operator-facing situational awareness by converting detected events into concise live summaries and context-grounded responses based on recent event history.
- Domain-adapted LLMs provide a plausible future path for improving technical consistency and microgrid-specific reasoning in the analyst layer.

Reference

F. Amjad, T. Korõtko, and A. Rosin, "Power-Q: A Hybrid Screener-Analyst Framework Using LLM Agents for Real-Time Grid Monitoring and Diagnostics," *IEEE Access*, vol. 14, pp. 61163–61177, 2026, doi: 10.1109/ACCESS.2026.3683539.

IEEE Recommended Practice for Monitoring Electric Power Quality', IEEE Std 1159-2019 Revis. IEEE Std 1159-2009, pp. 1–98, Aug. 2019, doi: 10.1109/IEEESTD.2019.8796486.