WHITEPAPER

Distributed, Power-Sovereign, AI Infrastructure for the Inference Economy

Executive Summary

AI is driving unprecedented growth in global compute demand, and traditional data-center models are hitting hard constraints in power, land, and latency. Electricity consumption from data centers is projected to rise sharply as AI workloads scale.[1] Meanwhile, grid interconnection queues are lengthening, suitable land near substations and fiber is scarce, and more workloads require real-time inference close to where data is generated.[2]

LT350 introduces a fundamentally different architecture: distributed, power-sovereign, modular AI data centers deployed in existing parking lots, integrating GPU cartridges, memory cartridges, solar generation, battery storage, and local fiber backhaul. This model directly addresses the structural constraints facing hyperscalers, enterprises, hospitals, and autonomous-vehicle fleets.

1. The AI Infrastructure Crunch

1.1 Explosive Growth in Data-Center Power Demand

Global data-center electricity consumption is expected to double or more over the next decade as AI training and inference workloads scale.[3] AI-optimized data centers are projected to grow even faster, with some analyses forecasting triple-digit percentage increases in power demand.[4]

---

[1] International Energy Agency (IEA), *Data Centres and Energy*, 2024.

[2] CBRE, North America Data Center Trends, 2024.

[3] Federal Energy Regulatory Commission (FERC), *Interconnection Queue Assessment*, 2024.

[4] McKinsey & Company, *AI and Power Demand*, 2024.

In the U.S., data centers could represent a high-single-digit share of total electricity consumption within a few years.[5]

1.2 Grid Interconnection Delays and Regional Strain

Interconnection queues for new generation and large loads have grown dramatically, with median wait times stretching into the multi-year range.[6] Only a minority of projects entering these queues ultimately reach commercial operation.

For data-center developers, this means:

- Long, uncertain timelines
- High upfront risk
- Exposure to shifting regulatory frameworks

These delays are already pushing operators toward behind-the-meter power, on-site generation, and distributed architectures.[7]

1.3 Land Scarcity and Siting Friction

Suitable land for traditional data centers is increasingly scarce in mature markets:

- Large parcels near substations and fiber
- Compatible zoning
- Community acceptance

Industrial land prices in key hubs have risen sharply, and competition from logistics and manufacturing has intensified.[8] Zoning, environmental review, and community opposition can add years to project timelines.

1.4 The Shift Toward Distributed Data Centers

AI inference is moving closer to where data is generated. Enterprises are deploying compact, GPU-accelerated clusters in hospitals, retail hubs, factories, and mobility depots.[9]

This shift favors distributed, high-performance nodes that deliver:

[5] U.S. Department of Energy's (DOE) Lawrence Berkeley National Laboratory (Berkeley Lab), *2024 Report on U.S. Data Center Energy Use*
[6] Federal Energy Regulatory Commission (FERC), *Interconnection Queue Assessment*, 2024.
[7] DataCenterDynamics, *Grid Constraints and AI Growth*, 2024.
[8] JLL, *Global Data Center Real Estate Report*, 2024.
[9] Wired, *The Rise of Distributed AI Compute*, 2024.

- Millisecond-level latency
- Local data sovereignty
- Local power solutions
- High availability independent of regional grid stress

LT350 is purpose-built for this emerging landscape.

2. LT350's Architecture: Modular AI Canopies in Existing Parking Lots

LT350 builds modular AI data-center canopies that install directly over existing parking lots without impacting parking. The ceiling of each canopy leverages the parking lot airspace and LT350's proprietary design to become a self-contained micro-datacenter by integrating with:

GPU Cartridges

- Each cartridge houses a small bundle of high-end GPUs
- Hot-swappable and incrementally deployable

Memory Cartridges

- Dedicated memory tiers for KV-cache offload
- Optimized for long-context and agentic inference workloads

Battery Cartridges

- Behind-the-meter storage
- Peak-shaving and curtailment resilience

Solar Integration

- Canopy-mounted solar generation
- Directly feeds local storage and compute loads

Local Fiber + Security

- Short-run, high-bandwidth connectivity
- Segmented networking for regulated workloads

Physical Isolation

- Dedicated infrastructure per customer
- Supports healthcare, financial, and defense-aligned deployments

This creates a power-sovereign, latency-optimized, modular inference node deployable where traditional data centers cannot go.

3. Structural Advantages of LT350's Model

3.1 Power Sovereignty via Solar + Battery

LT350's behind-the-meter architecture provides:

- Peak-shaving
- Curtailment resilience
- Predictable power cost
- Reduced interconnection burden

As regulators push large loads to "bring their own power," hybrid solar-plus-storage models are gaining traction [Source: BloombergNEF, 2024][10].

3.2 Near-Zero Land Cost and Rapid Deployment

Parking lots are:

- Already paved
- Already utility-connected
- Already zoned
- Widely distributed

This allows LT350 to:

- Avoid land acquisition
- Avoid greenfield construction
- Minimize permitting friction
- Deploy in weeks or months, not years

This dual-use approach mirrors trends in Europe, where solar canopies are being deployed over parking lots at scale.[11]

3.3 Modular, Demand-Matched GPU Deployment

Traditional data centers deploy multimegawatt increments, creating utilization risk. LT350's 3-GPU cartridge model enables:

- Small, revenue-aligned scaling

---

[10] BloombergNEF, *Hybrid Energy Systems for Compute*, 2024.
[11] Tech Trends Report (18th Edition), *Energy & Climate*, 2025.

- Rapid refresh cycles
- Reduced stranded capex

This keeps LT350's fleet closer to the performance frontier than hyperscale deployments.[12]

3.4 Proximity to High-Value, Regulated Customers

LT350 can deploy canopies within tens to hundreds of feet of:

- Hospitals
- Financial institutions
- Defense facilities
- Biotech campuses
- Autonomous-vehicle depots

This proximity enables:

- Deterministic low latency
- Physical data sovereignty
- Dedicated hardware
- Simplified compliance

These customers often pay premium pricing for these guarantees.[13]

3.5 Memory-Augmented Inference Nodes

As context windows expand and agentic workflows grow, inference bottlenecks shift to:

- KV-cache size
- Memory bandwidth
- Cross-GPU communication

LT350's memory cartridges:

- Offload KV-cache
- Increase throughput
- Support long-context workloads
- Improve economics for high-value inference

This positions LT350 as a specialized inference fabric, not just a GPU host[14].

---

[12] TechCrunch, *AI Hardware Deployment Models*, 2024.
[13] Deloitte, *Regulated Industries and Edge Compute*, 2024.
[14] LinkedIn Pulse article by SysHC, *The Era of Specialized Hardware: AI Inference and Video Encoding*, 2026

3.6 Alignment with Autonomous-Vehicle Infrastructure

Autonomous vehicles require:

- High-bandwidth data offload
- Frequent model updates
- Local inference
- Resilient power

LT350's canopies can be deployed in:

- Mall parking lots
- Airport structures
- Hospital campuses
- Fleet depots
- Convenience stores

This creates a distributed compute mesh aligned with AV fleet distribution.[15]

4. Market Alignment: Why LT350 Is Positioned for the Next Decade

4.1 Inference Is Moving to the Edge

Real-time AI requires:

- Millisecond latency
- Local data processing
- Local compliance
- Local power availability

Centralized cloud regions cannot meet all these constraints simultaneously.[16]

4.2 Power Scarcity Favors Behind-the-Meter Models

Data centers are becoming a major driver of electricity demand growth. Regulators are signaling that large loads must either:

- Pay more for grid upgrades, or
- Provide their own power flexibility

---

[15] CNBC, *Autonomous Vehicles and Infrastructure*, 2024.
[16] Wired, *The Rise of Distributed AI Compute*, 2024.

LT350's architecture aligns with this shift.[17]

4.3 Land Scarcity Favors Dual-Use Sites

Industrial land is scarce, expensive, and politically sensitive. Parking-lot deployment:

- Avoids land competition
- Avoids zoning battles
- Improves the underlying asset

This is a structural advantage over hyperscalers.[18]

5. Conclusion

LT350 is a new category of AI infrastructure:

- Distributed
- Power sovereign
- Memory augmented
- Latency optimized
- Modular
- Parking lot deployable

As AI shifts from centralized training to pervasive, real-time inference, LT350 believes it is positioned to become the default edge-compute fabric for hospitals, financial institutions, defense and biotech campuses, and autonomous vehicle fleets.

---

[17] BloombergNEF, *Hybrid Energy Systems for Compute*, 2024.
[18] JLL, *Global Data Center Real Estate Report*, 2024.